APE-Type Non-LTR Retrotransposons of Multicellular Organisms Encode Virus-Like 2A Oligopeptide Sequences, Which Mediate Translational Recoding during Protein Synthesis

Valerie Odon,[†,1] Garry A. Luke,[†,1] Claire Roulston,[†,1] Pablo de Felipe,[‡,1] Lin Ruan,[§,1] Helena Escuin-Ordinas,[¶,1] Jeremy D. Brown,[2] Martin D. Ryan,[*,1] and Andriy Sukhodub[1]

[1]Biomedical Sciences Research Complex, Biomolecular Sciences Building, University of St Andrews, Fife, United Kingdom

[2]RNA Biology Group and Institute for Cell and Molecular Biosciences, The Medical School, Newcastle University, Newcastle upon Tyne, United Kingdom

[†]These authors contributed equally to this work.

[‡]Present address: Spanish Medicines Agency (AEMPS), Parque Empresarial "Las Mercedes," Madrid, Spain

[§]Present address: Oakland Innovation Ltd., Cambridge Science Park, Cambridge, United Kingdom

[¶]Present address: Division of Hematology/Oncology, Department of Medicine, University of California at Los Angeles (UCLA), Los Angeles, CA

***Corresponding author:** E-mail: martin.ryan@st-and.ac.uk.

Associate editor: Howard Ochman

Abstract

2A oligopeptide sequences ("2As") mediate a cotranslational recoding event termed "ribosome skipping." Previously we demonstrated the activity of 2As (and "2A-like sequences") within a wide range of animal RNA virus genomes and non-long terminal repeat retrotransposons (non-LTRs) in the genomes of the unicellular organisms *Trypanosoma brucei* (*Ingi*) and *T. cruzi* (*L1Tc*). Here, we report the presence of 2A-like sequences in the genomes of a wide range of multicellular organisms and, as in the trypanosome genomes, within non-LTR retrotransposons (non-LTRs)—clustering in the *Rex1*, *Crack*, *L2*, *L2A*, and *CR1* clades, in addition to *Ingi*. These 2A-like sequences were tested for translational recoding activity, and highly active sequences were found within the *Rex1*, *L2*, *CR1*, and *Ingi* clades. The presence of 2A-like sequences within non-LTRs may not only represent a method of controlling protein biogenesis but also shows some correlation with such apurinic/apyrimidinic DNA endonuclease-type non-LTRs encoding one, rather than two, open reading frames (ORFs). Interestingly, such non-LTRs cluster with closely related elements lacking 2A-like recoding elements but retaining ORF1. Taken together, these observations suggest that acquisition of 2A-like translational recoding sequences may have played a role in the evolution of these elements.

Key words: **retrotransposon, 2A-like sequences, translational recoding, APE-type non-LTR.**

Introduction

The most ancient clades of non-long terminal repeat (LTR) retrotransposons (non-LTRs) within the *CRE*, *NeSL*, *R2*, *Hero*, and *R4* clades possess a single open reading frame (ORF) encoding a multifunctional protein comprising reverse transcriptase (RT) and restriction enzyme-like endonuclease (REL-endo) domains. One clade (*Dualen/RandI*) possesses an additional apurinic/apyrimidinic DNA endonuclease (APE) domain, thought to represent an intermediate stage leading to the evolution of a more advanced and diverse series of APE-type non-LTRs, in which the REL-endo domain was lost (reviewed in Malik et al. 1999; Kapitonov et al. 2009; Novikova and Blinov 2009). The 5′-region of the APE-type non-LTRs is, however, plastic in that many of these elements possess two ORFs (ORF1 and ORF2), whereas others lack an ORF1 (e.g., *L1Tc*, *Ingi*, and *BfCR1*; Albalat et al. 2003; Heras et al. 2006). Although coexpression of ORFs 1 and 2 in *cis* is essential for retrotransposition (Moran et al. 1996), bioinformatic analyses on different clades reveal a range of different ORF1 proteins suggesting that each type was acquired by independent evolutionary events. For simplicity, we will refer to the long ORF (encoding APE and RT domains) as "ORF2" throughout the text later, even though in some cases no ORF1 is present.

For non-LTRs with ORFs 1 and 2, both are encoded on a single transcript mRNA. The mechanism by which the second ORF is translated from the single polycistronic mRNA is, however, not clear (Alisch et al. 2006). In the case of the SART1 element, it has been shown that ORFs 1 and 2 are linked by an overlapping stop–start codon (-UAAUG-). The efficiency of the initiation of translation of ORF2 was shown to be dependent upon an RNA secondary structure downstream of this site: increasing the distance between the ORF1 stop codon/ORF2 start codon decreased the efficiency of the initiation of

translation of ORF2 (Kojima et al. 2005). This strategy of termination–reinitiation is also used by a variety of RNA viruses: influenza viruses (Horvath et al. 1990; Powell et al. 2008), respiratory syncytial viruses (Ahmadian et al. 2000; Gould and Easton 2005), pneumoviruses (Gould and Easton 2007), and caliciviruses (Meyers 2003, 2007; Luttermann and Meyers 2007).

Previously, we have reported the presence of "2A" translational recoding elements in the N-terminal region of the ORF2p of non-LTRs of *Trypanosoma cruzi* (*L1Tc*) and *T. brucei* (*Ingi*) (Heras et al. 2006). Such recoding elements are used in the genomes of many different RNA viruses (Donnelly et al. 1997; Donnelly, Hughes, et al. 2001; Luke et al. 2008): another relationship between the control of protein biogenesis in viruses and non-LTRs to parallel that of termination–reinitiation. These virus and non-LTR 2A oligopeptide sequences (2As) were shown to be active translation recoding elements by their insertion (in-frame) into an artificial polyprotein assay system (Donnelly, Hughes, et al. 2001; Donnelly, Luke, et al. 2001; Heras et al. 2006; Luke et al. 2008). Subsequent bioinformatic analyses showed 2As in the same region of non-LTRs of other trypanosome species (*T. vivax* and *T. congolense*; Heras et al. 2006).

"2A" derives from the systematic nomenclature of protein domains within the polyproteins of picornaviruses, a family of viruses with positive-stranded RNA genomes. 2As were first characterized in the central region of the foot-and-mouth disease virus (FMDV) polyprotein, between the upstream capsid and the downstream replication protein domains. 2A and "2A-like" oligopeptide sequences mediate a newly discovered form of translational recoding event termed variously as "ribosome skipping," "stop carry-on," or "stop-go" translation (Ryan et al. 1991, 1999; Ryan and Drew 1994; Donnelly et al. 1997; Donnelly, Hughes, et al. 2001; Donnelly, Luke, et al. 2001; de Felipe et al. 2003; Atkins et al. 2007; Doronina, de Felipe, et al. 2008; Doronina, Wu, et al. 2008; Brown and Ryan 2010; Sharma et al. 2012). Briefly, when a ribosome encounters 2A within an ORF, it "skips" the synthesis of a specific glycyl-prolyl peptide bond. The nascent protein is released from the ribosome by eukaryotic translation release factors 1 and 3 (eRF1/eRF3), thereby forming the C-terminus of 2A. Subsequently, ribosomes may then either terminate translation or resume translation of the downstream sequences as a discrete translation product. In this manner, multiple translation products are derived from a single ORF.

A motif at the C-terminus of 2A (-GD[V/I]ExNPG$^{\downarrow}$P-; "cleavage" site indicated by vertical arrow) is conserved among 2A-like sequences. Using this motif to probe databases revealed the presence of 2A-like sequences in a range of other mammalian, insect and crustacean RNA viruses. This motif alone does not, however, comprise an active 2A. The nature of the sequence immediately upstream of this motif, although not conserved among different 2A-like sequences, is critical for recoding activity (Ryan and Drew 1994; Sharma et al. 2012). Indeed, at that time, we detected a number of such motifs within cellular genes but only in the case of *L1Tc* and *Ingi* were the 2A-like sequences active in mediating translational recoding. As the range of cellular genome sequences has expanded, our recent bioinformatics analyses revealed the presence of 2A-like sequences within APE-type non-LTRs within the genomes of multicellular organisms: vertebrates, cephalochordates, molluscs, echinoderms, and cnidarians.

A number of factors support the notion that the acquisition of 2A-like sequences has played a role in the evolution of these APE-type non-LTR retrotransposons: 1) with a single exception, these 2A-like sequences all occur in the same N-terminal region of ORF2p, 2) their presence within a number of different non-LTR clades, 3) their presence within non-LTRs of a diverse range of species, and 4) that of the approximately 50 non-LTRs encoding 2A-like sequences we identified, the majority encode only one ORF.

Results

Identification of Non-LTRs Encoding 2A-Like Sequences

Probing databases with the 2A "signature" motif (-GD(V/I) ExNPGP-) revealed a number of non-LTRs encoding 2A-like sequences—from a range of species: *Xenopus tropicalis* (African claw-toed frog: vertebrate), *Branchiostoma floridae* (Amphioxus, Florida lancelet, cephalochordate), *Aplysia californica* (California sea slug, mollusc), *Crassostrea gigas* (Pacific oyster, mollusc), *Lottia gigantean* (Owl limpet, mollusc), *Strongylocentrotus purpuratus* (purple sea urchin, echinoderm), and *Nematostella vectensis* (sea anemone, cnidarian). Furthermore, the notion that these sequences have a biological role in the replication of such non-LTRs is supported by the observation that they are all located within approximately 40–80 aa from the N-terminus of ORF2; the same position as for the *L1Tc* and *Ingi* trypanosome 2A-like sequences (figs. 1*A* and 2*A*).

During the process of retrotransposition, non-LTRs may undergo truncation—to one degree or another—at their 5′-ends, such that the authentic ORF2 initiation codon may be deleted. If this is the case, bioinformatic algorithms then initiate translation from the next in-frame methionine codon, further truncating the protein sequence entered into the database. Because the 2A-like sequences (our database probe) are present within this 5′-region, this effect necessarily reduced our identification of such elements.

Phylogenetic Analyses

Those elements we identified that retained this 5′-region of ORF2, and that also possessed 2A-like sequences, clustered into six clades: *Rex1*, *L2*, *L2A*, *Crack*, *CR1*, and *Ingi* (fig. 1*B*). Note: A FASTA file and alignment of all RTclass1 domain sequences are supplied in the supplementary data, Supplementary Material online, together with a dendrogram file with bootstrap data. Non-LTRs encoding 2A-like sequences cluster alongside those with the "classical" organization observed for APE-type non-LTRs: An ORF1 (ORF1p comprising PHD/esterase domains) and ORF2 (ORF2p comprising apurinic/apyrimidinic endonuclease (APE) and RT domains. However, the majority of the approximately

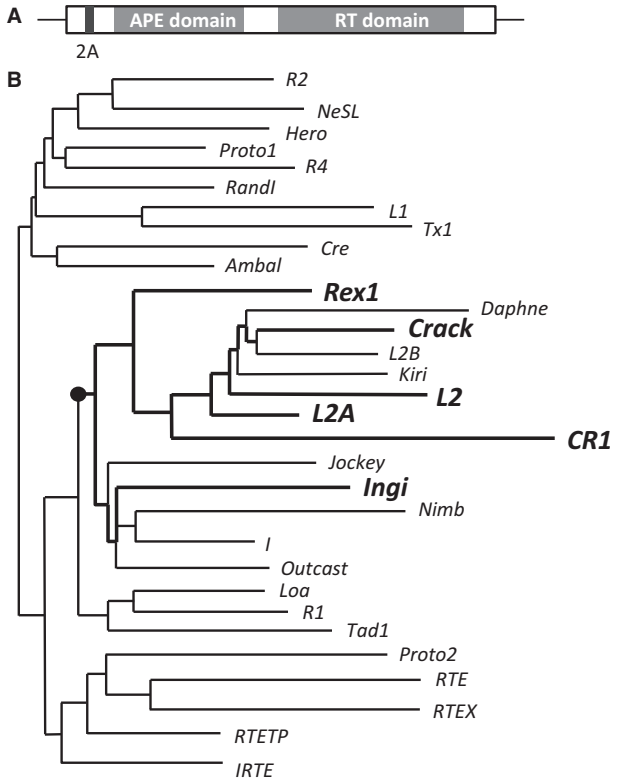

Fig. 1. 2A-like sequences within non-LTRs. The site of 2A-like sequences and the domain organization within ORF2 of "young" non-LTRs is shown (*A*). Dendrogram of non-LTR RT domains. The RTclass1 tree file was downloaded from Genetic Information Research Institute and adapted, such that each clade is represented by a single line, lengths representing the most unrelated element within each clade. Clades comprising non-LTRs encoding 2A-like sequences (active or inactive) are shown in bold and branches highlighted, the latest common ancestral node indicated with a closed circle (*B*).

50 elements encoding 2A-like sequences we identified appear to lack an ORF1.

The single exception to this pattern was the *L2A-1_NVe* non-LTR from *N. vectensis*. In this case, the 2A-like sequence was found within ORF1p. Alignments based upon the ORF2p RT domain of *L2A-1_NVe* show this element clusters within the L2A clade. Interestingly, we could not detect any 2A-like sequences within ORF2p of non-LTRs within the L2A clade.

Activity Assays

Although the -GD(V/I)ExNPGP- motif is conserved among 2A recoding elements, this tract is, by itself, not active; a suitable upstream context is required (Ryan and Drew 1994; Donnelly, Hughes, et al. 2001; Donnelly, Luke, et al. 2001; Doronina, de Felipe, et al. 2008; Doronina, Wu, et al. 2008; Sharma et al. 2012). One would expect, therefore, that not all 2A-like sequences identified by probing databases would be active in promoting translational recoding. To test for this activity, 2A-like sequences representative of the different non-LTR lineages were inserted into an artificial polyprotein system. This comprised a single, long, ORF encoding green fluorescent protein (GFP: stop codon removed), the 2A-like sequence to be tested, the β-glucuronidase (GUS: initiation codon removed), such that the single, long ORF was maintained. An inactive 2A would result in the single translation (fusion protein) product [GFP-2A-GUS]. An active 2A would produce the additional "cleavage" products of GFP with a C-terminal extension of 2A ([GFP-2A]), plus GUS. Not all 2A-like sequences are equally active: in our model of 2A-mediated translational recoding, the interaction of the nascent 2A with the ribosome exit tunnel determines the degree of accessibility of the peptidyl-tRNA ester linkage (in the P-site of the ribosome peptidyl-transferase centre) for the nucleophile—proly-tRNA (in the A site), hence the proportion of translational product in which the peptide bond is formed. Figure 2*A* shows those 2A-like sequences present in non-LTRs identified, arranged by clade. 2A-like sequences representative of each clade were inserted into the artificial polyprotein reporter system and the "cleavage" activity analyzed. Such activity analyses performed using translation systems in vitro have been shown to be reliable indicators of their activities within a range of (eukaryotic) cellular systems (Donnelly et al. 1997; Donnelly, Hughes, et al. 2001; Donnelly, Luke, et al. 2001; de Felipe et al. 2003; Doronina, de Felipe, et al. 2008; Doronina, Wu, et al. 2008; Luke et al. 2008).

In the case of the *Rex1* clade, we chose to analyze two sequences (*STR-61_SP* and *STR-197_SP*) with a substitution (E → D and E → N, respectively) at the same site within the canonical motif (-GD[V/I]ExNPG↓P-; fig. 2*A*). Both sequences were tested and shown to be active (fig. 2*B*). Although the uncleaved form ([GFP-2A-GUS]) was apparent, GUS and [GFP-2A] represented the major translation products. These data (plus those from other clades, see later) show that conservative changes at this site (E → Q/D/N) retain low activity (*STR-61_SP*, E → D; *STR-69_SP*, E → Q; *STR-197_SP*, E → N), whereas sequences with nonconservative substitutions at this site are not (fig. 2*A* and *B*).

In the case of the *L2* clade, *STR-51_SP* conformed to the motif, whereas *STR-69_SP* had a single substitution (E → Q: fig. 2*A*) at the same site as discussed earlier: both were active in mediating ribosome skipping (fig. 2*B*). Interestingly, mutation of this residue back to the canonical motif (*STR-69_SP*^{mut}; Q → E: fig. 2*A*) did not improve cleavage activity, in fact slightly more uncleaved, and slightly less cleavage products were observed. This single substitution (reconfirmed by additional nucleotide sequencing) produced a [GFP-2A] cleavage product, which migrated slightly more slowly than the wild-type counterpart (fig. 2*B*). Again, these data are consistent with our model of 2A-mediated "cleavage," in that the conserved motif alone is not sufficient for "cleavage": Interactions between the motif and upstream context (plus the upstream context and the ribosome exit tunnel) are essential for activity (Ryan and Drew 1994; Ryan et al. 1999; Donnelly, Luke, et al. 2001; Brown and Ryan 2010; Sharma et al. 2012).

In the *Crack* clade, *Crack-15_BF* and *Crack-17_BF* 2A-like sequences showed very low activity, only a small proportion of the radiolabel was present in the [GFP-2A] and GUS cleavage products: Both had a nonconservative substitution within the motif at the same site (E → H and E → A; fig. 2*A*). Indeed,

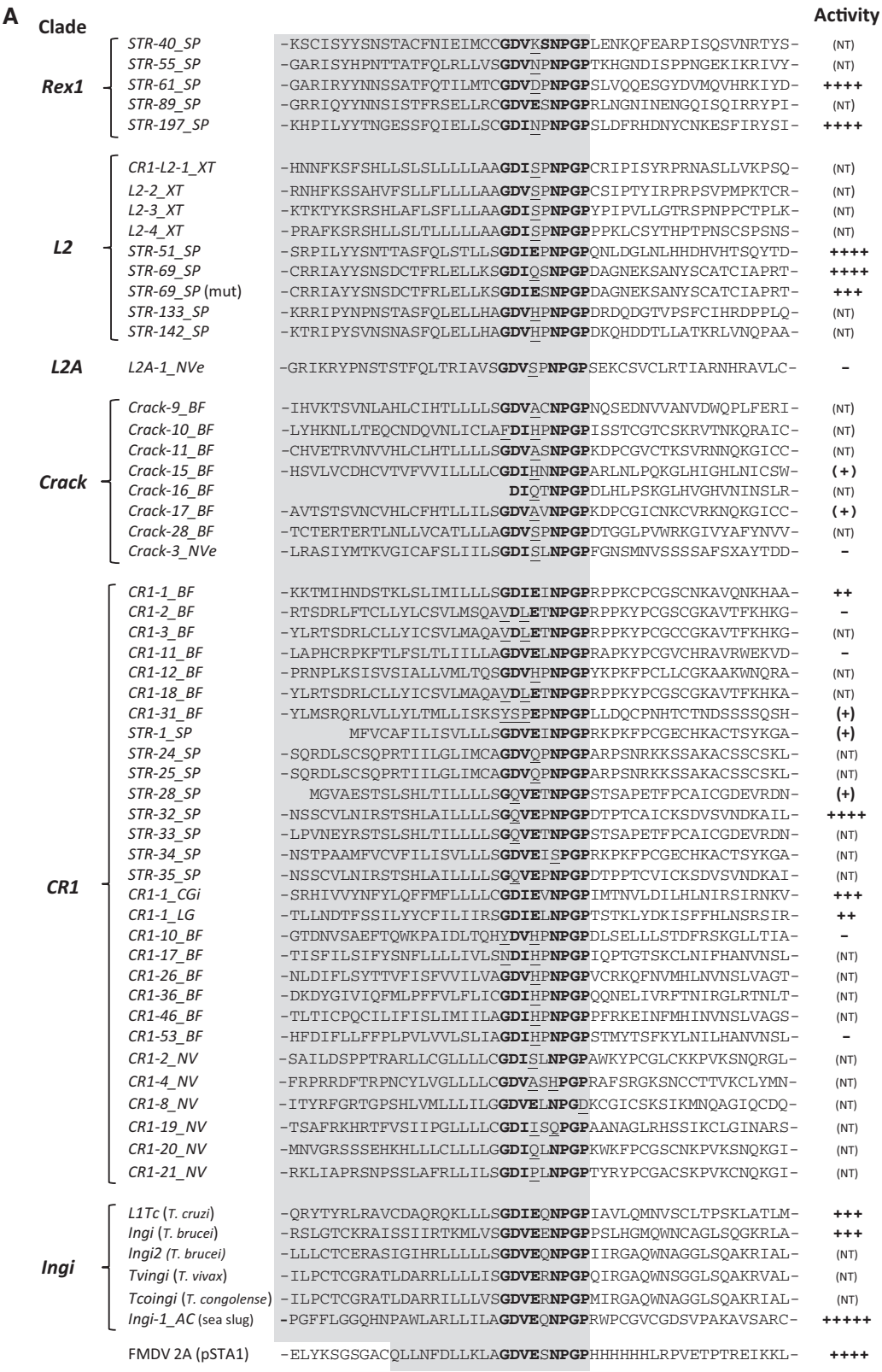

Fig. 2. 2A-like sequences and activity assays. 2A-like sequences of non-LTRs (plus the 20 aa downstream of the cleavage site) are shown together with FMDV 2A, for comparison. The 2A region is highlighted by the gray box. Residues conforming to the consensus motif are indicated in bold, those key residues which differ being underlined. Sequences are arranged by their order arising from sequence alignment (supplementary data, Supplementary Material online) (*A*). Coupled transcription/translation rabbit reticulocyte lysates were programmed with plasmid DNA as indicated (ordered as in *A*) and protein synthesis de novo monitored by the incorporation of ^{35}S-methionine. Translational recoding or "cleavage" activity was determined by the distribution of radiolabel within either the "uncleaved" form ([GFP-2A-GUS]) or the "cleavage" products ([GFP-2A] plus GUS) (*B*).

(continued)

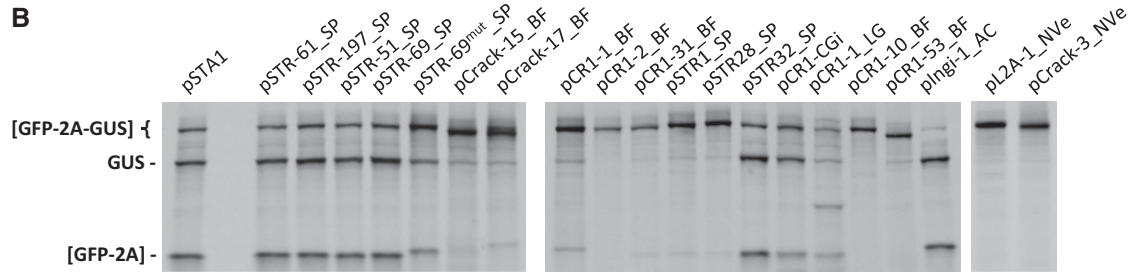

FIG. 2. Continued.

all the 2A-like sequences within this clade, both the *B. floridae* and the *N. vectensis* sequences, bore a change from the motif at this key site (fig. 2*A*), shown to be an important determinant of recoding activity by site-directed mutagenesis (Donnelly, Hughes, et al. 2001) and analyses of natural sequence variation (Luke et al. 2008).

For elements with the CR1 lineage, the recoding activity of 2A-like sequences was determined for five cephalochordate non-LTRs (*B. floridae*; *CR1-1_BF*, *CR1-2_BF*, *CR1-10_BF*, *CR1-31_BF*, and *CR1-53_BF*), three echinoderm (*S. purpuratus*: *STR-1_SP*, *STR-28_SP*, *STR-32_SP*), and two molluscs (*C. gigas*: *CR1-1_CGi*, *L. gigantean*: *CR1-1_LG*). Although the *Branchiostoma CR1-1_BF* 2A-like sequence showed low recoding activity, the others showed extremely low (*CR1-31_BF*) activity, or no activity—essentially too low for us to detect using this system (*CR1-2_BF*, *CR1-10_BF*, and *CR1-53_BF*; fig. 2*B*). Although the canonical motif was largely conserved in these sequences, we have shown that this motif must have an appropriate tract immediately upstream to mediate ribosome skipping (Ryan and Drew 1994; Ryan et al. 1999; Donnelly, Luke, et al. 2001; Brown and Ryan 2010; Sharma et al. 2012). In the case of the echinoderm CR1 elements, both *STR-1_SP and STR-28_SP* are N-terminally truncated forms and both showed very low levels of activity. The 2A-like sequence of *STR-1_SP* conforms to the motif, whereas *STR-28_SP* has a single substitution at a residue (D → Q: fig. 2*A*), whose identity was shown to be an important determinant of recoding activity (Donnelly, Hughes, et al. 2001). This same substitution is observed, however, within the active *STR-32_SP* 2A-like sequence: The truncation or substitutions (compared with *STR-32_SP*) within the tract immediately upstream of the motif in *STR-1_SP* and *STR-28_SP* renders these 2A-like sequences largely inactive. In both of the mollusc *CR1* elements (*CR1-1_CGi* and *CR1-1_LG*), the 2A-like sequences were active. In the case of *CR1-1_LG*, an additional translation product was generated from internal initiation of translation, a common feature of translation reactions in vitro.

Previously, we had tested 2A-like sequences from *L1Tc* and *Ingi* elements from trypanosome species and found these were active. Here, we show that the 2A-like sequence within the *Ingi-1_AC* non-LTR of the California sea slug (*A. californica*: mollusc) is highly active. A very high proportion of the radiolabel (>97%) is present in the [GFP-2A] and GUS "cleavage" products, with only a very small proportion in the "uncleaved" [GFP-2A-GUS] form (fig. 2*B*). Indeed, this 2A-like sequence is more active than the virus (FMDV) sequence.

2A-Like Sequences within *N. vectensis* Non-LTRs

Interestingly, a series of non-LTRs within the genome of the sea anemone *N. vectensis* (Putnam et al. 2007) also encode 2A-like sequences, clustering within the *CR1* and *Crack* clades (*CR1-2/4/8/19/20/21_NV*; *Crack-3_NVe*; fig. 2*A*). Again, each of these 2A-like sequences is found in the N-terminal region of ORF2, as observed for all other 2A-like sequences discussed earlier (fig. 1*A*). In all cases, however, mutations are observed at a key residue(s) within the canonical motif; *CR1-2_NV* = E → S, *CR1-4_NV* = E → A, *CR1-8_NV* = PGP → PGD, *CR1-19_NV* = E → I, *CR1-20_NV* = ES → QL, *CR1-21_NV* = ES → PL and *Crack-3_NVe* = E → S (fig. 2*A*). Our previous site-directed mutagenesis analyses of 2A showed that such mutations ablate recoding activity, with the exception of the single point mutation of E → Q (Donnelly, Hughes, et al. 2001; *vide supra*).

A 2A-like sequence is also observed, however, in an element within the *N. vectensis* genome clustering within the *L2A* clade (*L2A-1_NVe*: fig. 2*A*). In this case, the 2A-like sequence is encoded within ORF1, rather than ORF2, as in all other cases. The ORF1 of *L2A-1_NVe* is 656 aa long, and, again, the 2A-like sequence is found in the N-terminal region (aa 72–102). As for the 2A-like sequences within the CR1 ORF2s, the 2A-like sequence in *L2A-1_NVe* encodes a serine at the site corresponding to the key glutamate residue (E → S; fig. 2*A*), discussed earlier. Both the *L2A-1_NVe* and *Crack-3_NVe* 2A-like sequences were tested and found to be inactive (fig. 2*B*).

Correlation of a Single ORF and the Presence of 2A

We have shown non-LTRs within five clades, which encode active 2A-like sequences, expanding the range of such non-LTRs from the single report for kinetoplastid genomes (*T. cruzi*, *T. brucei*, *T. vivax*, and *T. congolense*; Heras et al. 2006). During the course of our bioinformatic analyses, we noticed that the majority of non-LTRs encoding a 2A recoding element did not possess an ORF1, exceptions including *CR1-26_BF*, *CR1-53_BF*, and *CR1-1_LG*. As noted earlier, non-LTRs may undergo truncation of their 5′-ends during retrotransposition, such that ORF1 could be deleted entirely, although the high proportion of non-LTRs encoding a 2A-like sequence but lacking an ORF1 argues against this being purely an artifact. Sequences of non-LTRs encoding a 2A-like

sequence (lacking ORF1) cluster alongside elements from other species, which encode an ORF1—but not a 2A-like sequence. All elements within the *Ingi* clade (*Ingi*, *Tcoingi*, *Tvingi*, and *L1Tc*) do not appear to encode an ORF1 but possess 2A-like sequences (fig. 2*A*). It has been reported that *Vingi* non-LTRs only encode a single, long (993 aa), ORF (Kojima et al. 2011), but many of these elements appear to have undergone N-terminal truncation, and we were unable to detect any 2A-like sequences.

Non-LTR Genome Organization

In all but one of the elements present within the different clades, the site of insertion of the 2A-like sequence was the same: the N-terminal region of ORF2p. The single exception was the 2A-like sequence within the N-terminal region of ORF1p (*L2A-1_NVe*). In some cases, 2A-like sequences are present immediately upstream of the APE domain (e.g., *L1Tc*, *Ingi*, *CR1-17_BF*, *CR1-26_BF*, and *STR-194_SP*). In other cases, a PHD domain, observed within ORF1p of some non-LTRs, is found between the 2A-like sequence and the APE domain in ORF2p (e.g., *Crack-17_BF*, *CR1-1/17/26_BF*, and *STR-24/25/34/35_SP*). In the remaining cases, a tract of some 90–115 aa is present between the 2A-like sequence and the APE domain (e.g., *STR-51/61/142/197_SP*), with no motifs suggesting a function.

2A Recoding Activities

A previous study of 2A-like sequences in the genome of *T. cruzi* showed all *L1Tc* elements encoded a 2A-like sequence, although sequence heterogeneity was observed (Heras et al. 2006). The majority of elements (~57.5%) encoded the canonical 2A motif -DIEQNPGP-, whereas 20% of elements encoded a single N → H substitution within the motif (-DIEQHPGP-). Previously, this mutation (within FMDV 2A) had been created and shown to reduce "cleavage" activity (Donnelly, Hughes, et al. 2001). A similar effect was observed for the *L1Tc* 2A-like sequences (Heras et al. 2006). For non-LTRs encoding 2A-like sequences in the *Rex1*, *L2*, *Crack*, and *CR1* clades, frequent substitutions are observed at the glutamate residue (-GD(V/I)ExNPGP-: fig. 2*A*), previously identified by site-directed mutagenesis as an important determinant of "cleavage" activity (Donnelly, Hughes, et al. 2001). The CR1 clade has high heterogeneity at this residue and it may be that either 1) only very low levels of recoding activity is required from these particular 2A-like sequences or 2) previously (more) active 2A-like sequences have been rendered essentially inactive by the accumulation of mutations in such a key residue.

Discussion

Previously we have identified and characterized 2A translational recoding sequences in a wide range of mammalian, insect, and crustacean RNA virus genomes (Luke et al. 2008), plus non-LTR elements within the genome of unicellular organisms (trypanosomes; Heras et al. 2006). In this article, we provide the first evidence of active 2A-like sequences within the genomes of multicellular organisms: vertebrates, cephalochordates, molluscs, cnidarians, and echinoderms. 2A and 2A-like sequences have been widely used in biotechnology and have been shown to function in all eukaryotic systems tested to date (e.g., plant, fungal, yeast, insect, and mammalian cells), a reflection of the very high degree of conservation of the structure of the eukaryotic ribosome. It should be noted, however, that we have tested 2A-like sequences from a range of species in a single mammalian (rabbit)-derived cell-free translation system. Furthermore, our analyses are based upon the distribution of radiolabel in sodium dodecyl sulphate (SDS) gels by exposure to film, and it may be that our methods simply cannot physically detect the lowest levels of translational recoding activities, which still retain a biological activity within the organism in question.

Acquisition of 2A-Like Sequences

It is possible that the transfer of 2A-like sequences could be mediated by viruses: Active 2A-like sequences are present in the genome of viruses, which infect fish or crustaceans (Luke et al. 2008). Virus particles (or virus-like particles [VLPs]) can, however, encapsidate host-cell, rather than virus, RNAs. The RNA content of highly purified preparations of purified flock house virus (FHV), a nonenveloped RNA virus, and VLPs of FHV and the related *Nudaurelia capensis* omega virus were studied. In the case of VLPs, 5.3% of the packaged RNAs were found to be transposable elements derived from the host-cell genome. Authentic FHV virions also packaged a variety of host RNAs, including significant quantities of transposable elements (Routh et al. 2012). Naturally, packaging of these host non-LTRs into virus particles (which could deliver these genetic elements into the cytoplasm of cells of other species) constitutes a possible mechanism of horizontal sequence transfer. Neoplastic cells release an abundance of microvesicles, which have been shown to contain RNAs, including notably high levels of retrotransposon RNA transcripts (Balaj et al. 2011). Such microvesicles could provide another mechanism for horizontal sequence transfer via predation/ingestion and fusion of prey-derived microvesicles with cells of the predator delivering the nucleic acid into the cytoplasm. Indeed, "simple" host–parasite interactions are thought to play a role in horizontal transfer of transposons across phyla (Gilbert et al. 2010). Such events would need to occur either by transfer/integration into the genome of a totipotent somatic cell or into the genome of germ-line cells by either direct or indirect transfer (initial transfer into a somatic cell plus subsequent transfer to a germ-line cell by virus particles/microvesicles).

The 2A-like sequences we have detected all occur (except *L2A-1_NVe*) in the same (N-terminal) region of ORF2, suggesting a functional significance. The *Rex1* clade comprises non-LTRs from a wide range of species, yet all the occurrences of 2A-like sequences within this clade occur with the genome of a single species, *S. purpuratus* (echinoderm). The *Crack* clade comprises non-LTRs from a wide range of species, but the occurrences of 2A-like sequences within this clade occur only within the genomes of two species, *B. floridae*

(cephalochordate) and *N. vectensis* (cnidarian). The *L2*, *CR1*, and *Ingi* clades each comprise non-LTRs from a wide range of species, and in these cases, we observe sequences in the genomes of organisms, which diverged at an early stage in the evolution of metazoans; *L2* clade: *X. tropicalis* (vertebrate), *N. vectensis* (cnidarian), and *S. purpuratus* (echinoderm); the *CR1* clade: *B. floridae* (cephalochordate), *S. purpuratus*, *C. gigas* (mollusc), and *L. gigantean* (mollusc), and the *Ingi* clade: *T. brucei*, *T. cruzi*, *T. vivax*, *T. conglenese* (kinetoplastid), and *A. californica* (mollusc).

The Functions of Virus and Non-LTR 2A-Like Sequences

In general, virus 2A-like sequences are highly active and serve to bring about the rapid, cotranslational, separation of polyprotein domains. Such domains are synthesized as discrete translation products even though they are encoded by the same ORF. Some virus 2A sequences have evolved to produce a mixture of "cleaved" and uncleaved (fusion protein) translation products (Luke et al. 2008). Other virus 2A sequences appear to have been used, such that the genome has acquired new functions by essentially "bolting-on" an extra domain to an existing protein (extending the ORF), using 2A as a "linker" sequence. This is most clearly seen in the comparison of type A, B, and C rotaviruses, where type C viruses (but not type A or B) have an RNA binding domain linked to C-terminus of protein NS3 via a 2A linker (Luke et al. 2008). A similar extension is seen at the N-terminus of the long polyprotein encoded by the double-stranded RNA virus penaeid shrimp infectious myonecrosis virus (Luke et al. 2008). Why and from where such additional domains have arisen is not known, but there is evidence to support the case that 2A can be used to mediate the transposition of function between genetic elements. Indeed, 2A-like sequences are very widely used in animal/plant biotechnologies and biomedical applications for linking multiple functions into (mono-cistronic) "self-processing" polyproteins (http://www.st-andrews.ac.uk/ryanlab/page10.htm, last accessed June 13, 2013).

Although we have shown a range of translational recoding activities associated with these non-LTR 2A-like sequences, questions naturally arise as to their function (whether in ORF1p or ORF2p) and if such elements lacking ORF1 still retain their autonomy with regards retrotransposition. We have proposed that the 2A recoding element is able to downregulate the level of the translation product downstream of 2A compared with that upstream (Brown and Ryan 2010)—a translational regulatory element. With regards the retrotransposition of non-LTRs encoding active 2A-like sequences, two aspects arise from this activity. First, for optimal retrotransposition activity, an excess of the function encoded by sequences upstream of 2A is required over the functions encoded downstream of 2A (RT/APendonuclease). Cells employ a range of mechanisms to inhibit retrotransposition: Although it is thought certain evolutionary advantages may accrue from this activity, presumably too high a level of retrotransposition is disadvantageous. Because selection for, and maintenance of, non-LTRs depends entirely upon the "host" cell, it may be to the advantage of such elements to evolve mechanisms of "self-restraint" with regards the level of retrotransposition within the cell. It was proposed that the L1Tc 2A-like sequence may produce a downregulation of the non-LTR translational products downstream of 2A (APE and RT domains). This could help explain the observation that even though relatively high levels of L1Tc mRNAs are detected within cells, only low levels of ORF2p protein are detected (Heras et al. 2006).

Another important aspect of 2A-mediated "ribosome skipping" is that this activity produces discrete, but different, translation products from a single ORF. The ORF2 of APE-type non-LTRs is a multifunctional protein, yet the large majority of these elements encode other functions within a separate ORF1 and not fused to the ORF2 multifunctional protein. Here, one may draw an analogy with RNA replication and polyprotein processing in positive-stranded RNA viruses. During the replication of such virus genomes, some proteins have an obligate function in *cis* (acting upon the very same RNA molecule on which they are encoded), but may also function *in trans*. Other virus proteins (notably capsid proteins) function *in trans* and are generated (or separated from replication proteins) by a variety of methods in different virus groups: 1) a rapid, cotranslational, "cleavage" of the polyprotein (e.g., picornaviruses), 2) by being encoded in a separate ORF within the single-stranded genomic RNA (e.g., dicistroviruses), 3) by being encoded in a separate ORF(s) on subgenomic RNA transcripts produced from a genome-length RNA template (e.g., coronaviruses), or 4) by being encoded by a separate genomic RNA strand altogether (e.g., comoviruses). Drawing upon this analogy with the replication strategy of positive-stranded RNA viruses, one could argue that non-LTR ORF2 functions have an obligate function *in cis* (reverse transcription/integration into the genome) but can also function *in trans* (e.g., SINE transposition by LINE-encoded functions). For ORF1 functions, however, the non-LTR genome organization (ORF1 + ORF2) suggests that ORF1 (functions) need to be generated as a translation product quite separate from the ORF2 multifunctional protein. Implicit in this argument is that encoding a 2A-like translational recoding sequence may have allowed APE-type non-LTR genome reorganization from ORF1 + ORF2 to a single ORF: Functions N-terminal of 2A may be generated in the form of a discrete translation product quite separate from the canonical ORF2 functions.

As mentioned earlier, the 2A-like sequences we have detected occur both in 1) different non-LTR clades and 2) a wide range of species. In all cases, they occur (except *L2A-1_NVe*) in the same N-terminal region of ORF2. This complete conservation of the site of 2A with ORF2 is highly suggestive of conserved function. In trypanosome genomes, the 2A-like sequences within L1Tc show a range of mutations, each with different "cleavage" activity (Heras et al. 2006). Similarly, in this article, we describe 2A-like sequences with a range of activities/no activity within the same species (e.g., *S. purpuratus* and *B. floridae*). The simplest explanation of these data is that during evolution, non-LTRs with active 2A-like sequences were acquired, but have subsequently undergone accumulation of mutations leading to a reduction/

loss of activity. An alternative explanation is that during evolution, a common progenitor form of these 2A-like sequences (recoding inactive) has undergone a series of independent mutations to produce the range of activities we report here.

We did not detect any non-LTR 2A-like sequences in the genomes of mammals, reptiles, birds, or fish. The *CR1-1_Bf* 2A-like sequence was the most active from a cephalochordate (*B. floridae*) genome. Given the limited genome data currently available, it is difficult to discern any pattern of distribution. As it stands, however, the distribution of 2A-like sequences we observed in non-LTRs is consistent with the model of deuterosome evolution proposed by Delsuc et al. (2006), in which a lineage comprising echinoderms and cephalochordates diverged from a lineage comprising tunicates and vertebrates. Analyses of complete genome sequences of the sea urchin (Sodergren et al. 2006), sea anemone (Putnam et al. 2007), and amphioxus (Putnam et al. 2008) led, however, to an evolutionary scheme in which the cephalochordates represent the most basal members of the chordate lineage, with tunicates forming a parallel "sister" lineage (Putnam et al. 2008). In this scheme, amphioxus (encoding mainly inactive 2A-like sequences) represents the most "basal" extent of an organism with a genome comprising non-LTRs encoding 2A-like sequences within the chordate lineage. In this case, the pattern of distribution of non-LTRs encoding 2A-like sequences within individual clades does, however, argue either for acquisition of 2A-like sequences within a very early ancestral form of non-LTR accompanied by a subsequent complex pattern of sequence loss.

An alternative model would be that 2A-like sequences were acquired by non-LTRs at a later stage in their evolution. However, because 2A-like sequences are found within a number of different clades of non-LTRs, this model invokes either a series of independent acquisitions or transfer of sequences between non-LTR in different clades: Possibly some aspect of the biology/molecular biology of these types of metazoan engenders a higher rate of horizontal sequence transfer. In the case of virus 2A-like sequences we have proposed a model of multiple, independent, acquisitions (Luke et al. 2008).

To date, genome sequences are available for only a very few organisms in the phyla/subphyla involved in this study. Interpretation of the pattern of the distribution of non-LTRs encoding 2A-like sequences—both in terms of the type (clade) of non-LTR and the species in which they occur, will undoubtedly change and become clearer as more genome sequences are determined, of the organisms themselves and the viruses which infect them. The occurrence of 2A-like sequences in non-LTRs represents, however, another fascinating parallel between virus genomes and non-LTR retrotransposons.

Materials and Methods

Database Probing

A "canonical" motif (-GD(V/I)ExNPGP-), derived from the comparison of conservation within different virus 2A-like sequences, was used to probe genome sequence databases maintained at the NCBI (http://www.ncbi.nlm.nih.gov/, last accessed June 13, 2013), the Pasteur Institute (Mobyle@pasteur http://mobyle.pasteur.fr/, last accessed June 13, 2013), Worm Base (http://www.wormbase.org/db/searches/blast_blat, last accessed June 13, 2013), the Max Planck Institute for Molecular Genetics (http://goblet.molgen.mpg.de/cgi-bin/seaurchin-genombase.cgi, last accessed June 13, 2013), the JGI Genome Portal (http://genome.jgi-psf.org/, last accessed June 13, 2013), UniProt (http://www.uniprot.org/, last accessed June 13, 2013), ENSEMBL Genomes (http://www.ensemblgenomes.org/, last accessed June 13, 2013), FlyBase (flybase.org/blast/), TriTrypDB (tritrypdb.org), HMMER (http://hmmer.janelia.org/search/phmmer, last accessed June 13, 2013), the Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/, last accessed June 13, 2013), ScanProsite (http://prosite.expasy.org/scanprosite, last accessed June 13, 2013), and REPBASE at the Genetic Information Research Institute (http://www.girinst.org/repbase/, last accessed June 13, 2013; Jurka et al. 2005).

Sequences Used in Bioinformatic Analyses

Non-LTRs are now designated by the clade/element name, underscore, then species identifier. Hence, *_AC* refers to *A. californica* (California sea slug: mollusc), *_BF—B. floridae* (Amphioxus: Florida lancelet: cephalochordate), *_CGi—C. gigas* (Pacific oyster: mollusc), *_HM—Hydra magnipapillata* (fresh water polyp: cnidarian), *_LG—L. gigantean* (owl limpet: mollusc), *_NV* (and *_NVe*)*—N. vectensis* (sea anemone: cnidarian), *_SP—S. purpuratus* (purple sea urchin: echinoderm), *_TB—T. brucei* (kinetoplastid), and *_XT—X. tropicalis* (African claw toed frog: chordate). Further information may be obtained from REPBASE. Genome and protein data were downloaded from the sites listed earlier.

We arbitrarily designated non-LTRs from *S. purpuratus* with "STR" identifiers: *STR-1_SP* (XP_797143), *STR-24_SP* (XP_001196407), *STR-28_SP* (XP_001179204), *STR-29_SP* (XP_791376), *STR-30_SP* (XP_001199602), *STR-31_SP* (XP_001200060), *STR-32_SP* (XP_001185404), *STR-33_SP* (XP_001184905), *STR-34_SP* (XP_001196844), *STR-35_SP* (XP_001200466), *STR-181_SP* (XP_001196407), *STR-51_SP* (GLEAN3_22449), *STR-69_SP* (GLEAN3_27016), *STR-133_SP* (GLEAN3_00868), *STR-142_SP* (GLEAN3_14631), and *STR-194_SP* (GLEAN3_18278).

Bioinformatic Analyses

Classification of non-LTRs was conducted using the Repbase RTclass1 web server (http://www.girinst.org/RTphylogeny/RTclass1/, last accessed June 13, 2013). The RT domains of all sequences used at GIRINST were downloaded and used to define this domain in non-LTRs encoding 2A-like sequences by a process of reiterative alignment using Muscle either locally (Unipro UGENE 1.11) or using a web-based algorithm (http://www.ebi.ac.uk/Tools/msa/muscle/, last accessed June 13, 2013), together with "trimming" to produce the alignment shown in the supplementary data, Supplementary Material online.

Table 1. Oligonucleotide Primer Sequences (Reversed, Complemented) that Encode 2A-Like Sequences Forming In-Frame Insertions between GFP and GUS: for Clarity, the 20 (5′) Nucleotides Complementary to GFP Are Omitted.

Non-LTR Designation	2A-Like Sequence
FMDV 2A (pSTA1)	Q L L N F D L L K L A **G D V E** S **N P G P** CAGCTGTTGAATTTTGACCTTCTTAAGCTTGCGGGAGACGTCGAGTCCAACCCCGGGCCC
STR-1_SP	M F V C A F I L I S V L L L S **G D V E I N P G P** ATGTTTGTGTGCGCGTTCATACTGATATCAGTATTGCTACTGAGTGGTGATGTGGAAATAAATCCCGGGCCC
STR-28_SP	M G V A E S T S L S H L T I L L L L S **G** Q **V E** T **N P G P** ATGGGTGTAGCTGAGTCGACTTCTTTGAGCCACCTAACCATCCTACTTCTCCTCAGCGGACAAGTTGAAACCAACCCTGGGCCC
STR-32_SP	N S S C V L N I R S T S H L A I L L L L S **G** Q **V E** P **N P G P** AACTCTTCATGTGTCCTCAACATTCGTTCCACCAGCCACCTGGCCATCTTACTACTTCTCAGTGGGCAAGTTGAGCCCAACCCAGGGCCC
STR-51_SP	S R P I L Y Y S N T T A S F Q L S T L L S **G D I E** P **N P G P** AGTAGACCAATATTGTATTATAGTAATACTACAGCAAGTTTCCAATTGAGTACCTTACTCTCTGGCGATATTGAGCCTAACCCAGGGCCC
STR-61_SP	G A R I R Y Y N N S S A T F Q T I L M T C **G D V** D P **N P G P** GGAGCCCGCATAAGGTATTACAATAACTCTTCTGCAACTTTTCAAACTATTCTTATGACCTGTGGAGATGTTGATCCCAACCCGGGGCCC
STR-69_SP	C R R I A Y Y S N S D C T F R L E L L K S **G D I** Q S **N P G P** TGTAGAAGAATTGCATACTACAGCAACAGTGACTGCACATTTAGGTTAGAACTTTTGAAATCAGGCGATATTCAATCTAACCCTGGGCCC
STR69^{mut}_SP	C R R I A Y Y S N S D C T F R L E L L K S **G D I E** S **N P G P** TGTAGAAGAATTGCATACTACAGCAACAGTGACTGCACATTTAGGTTAGAACTTTTGAAATCAGGCGATATTGAATCTAACCCTGGTCCT
STR-197_SP	K H P I L Y Y T N G E S S F Q I E L L S C **G D I** N P **N P G P** AAGCATCCAATACTTTACTATACCAATGGCGAGTCTTCCTTCCAGATTGAACTCCTTTCATGTGGTGATAATCAACCCCAACCCGGGCCC
CR1-1_BF	K K T M I H N D S T K L S L I M I L L L S **G D I E** I **N P G P** AAGAAAACAATGATTCACAATGATAGTACAAAGTTGTCACTGATTATGATCTTGCTCCTAAGTGGAGATATTGAGATCAACCCAGGGCCC
CR1-2_BF	R T S D R L F T C L L Y L C S V L M S Q A V **D** L **E** T **N P G P** CGAACATCAGACCGACTATTCACATGCCTACTATACCTATGCTCAGTACTAATGTCACAAGCAGTAGACCTAGAAACAAACCCCGGGCCC
CR1-10_BF	G T D N V S A E F T Q W K P A I D L T Q H Y **D V** H P **N P G P** GGAACAGACAACGTATCAGCAGAATTCACACAATGGAAACCAGCAATTGACCTAACACAACACTACGACGTACACCCAAACCCAGGGCCC
CR1-31_BF	Y L M S R Q R L V L L Y L T M L L I S K S Y S P **E** P **N P G P** TACCTAATGTCACGACAACGACTAGTACTACTATACCTAACAATGCTACTAATTAGCAAATCATACTCACCAGAACCAAACCCCGGGCCC
CR1-53_BF	H F D I F L L F F P L P V L V V L S L I A **G D I** H P **N P G P** CACTTCGACATTTTCCTACTATTCTTCCCACTACCAGTACTAGTAGTACTATCACTAATTGCAGGAGACATTCACCCAAACCCAGGGCCC
Crack-3_NVe	S I Y M T K V G I C A F S L I I L S **G D I** S L **N P G P** AGCATTTATATGACTAAAGTAGGTATTTGTGCATTTAGTCTTATTATTCTGAGTGGAGATATTAGTCTGAACCCTGGGCCC
Crack-15_BF	H S V L V C D H C V T V F V F I L L L L C **G D I** H N **N P G P** CACTCAGTACTAGTATGCGACCACTGCGTAACAGTATTCGTAGTAATTCTACTACTACTATGCGGAGACATTCACAACAACCCCGGGCCC
Crack-17-BF	A V T S T S V N C V H L C F H T L L I L S **G D V** A V **N P G P** GCAGTAACATCAACATCAGTAAACTGCGTACACCTATGCTTCCACACACTACTAATTCTATCAGGAGACGTAGCAGTAAACCCAGGGCCC
Ingi-1_AC	F L G G Q H N P A W L A R L L I L A **G D V E** Q **N P G P** TTTCTAGGTGGACAGCACAATCCAGCATGGCTAGCACGACTACTAATACTAGCAGGAGACGTAGAACAGAATCCAGGGCCC
CR1-1_CGi	S R H I V V Y N F Y L Q F F M F L L L L C **G D I E** V **N P G P** TCTAGACATATCGTAGTGTATAACTTCTATCTTCAATTCTTTATGTTCCTACTGCTACTCTGCGGAGACATAGAAGTAAATCCAGGGCCC
CR1-1_LG	N D T F S S I L Y Y C F I L I I R S **G D I E** L **N P G P** AACGACACATTCTCATCAATACTGTACTACTGCTTCATACTAATAATACGATCAGGAGACATAGAACTAAACCCAGGGCCC
L2A-1_NVe	K R Y P N S T S T F Q L T R I A V S **G D V** S P **N P G P** AAACGATATCCTAATAGTACAAGTACATTTCAACTAACACGAATTGCAGTTAGTGGAGATGTTCAGTCCAAATCCTGGGCCC

NOTE.—Residues conforming to canonical motif (-GD[V/I]ExNPGP-) are in bold, and those not conforming are underlined.

Cloning of 2A-Like Sequences

Sequences encoding 2A-like sequences were inserted in between GFP and GUS (plasmid pSTA1; Luke et al. 2008), such that the single ORF was maintained (table 1). The T7 forward primer was used to amplify GFP from pSTA1 (Donnelly, Hughes, et al. 2001; Donnelly, Luke, et al. 2001), whereas oligonucleotides encoding 2A-like sequences (together with 18 bases complementary to the 3′-end of GFP) were used as reverse primers. Polymerase chain reaction products were cloned into pGEM-T Easy (Promega), inserts excised with BamHI and ApaI, purified following agarose gel electrophoresis then ligated into pSTA1, similarly restricted. All plasmids were constructed using standard methods and confirmed by DNA sequencing.

Coupled Transcription/Translation In Vitro

Plasmids encoding 2A-like sequences were used to program a TNT Quick coupled transcription/translation System, according to the manufacturer's instructions (Promega). Protein synthesis de novo was monitored by the incorporation of ^{35}S-methionine and the distribution of radiolabel determined by SDS-polyacrylamide gel electrophoresis (PAGE) as described (Donnelly, Hughes, et al. 2001; Donnelly, Luke, et al. 2001). Briefly, 0.1 μg plasmid (1.0 μl) was mixed with 3 μCi ^{35}S-Met and 10 μl TNT T7 Quick Master Mix and incubated for 90 min at 30 °C in a 12.5 μl reaction volume. Translation products were then analyzed by SDS-PAGE (10%) and autoradiography.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC).

References

Ahmadian G, Randhawa JS, Easton AJ. 2000. Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism. *EMBO J.* 19:2681–2689.

Albalat R, Permanyer J, Cañestro C, Martínez-Mir A, Gonzàlez-Angulo O, Gonzàlez-Duarte R. 2003. The first non-LTR retrotransposon characterised in the cephalochordate amphioxus, BfCR1, shows similarities to CR1-like elements. *Cell Mol Life Sci.* 60:803–809.

Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV. 2006. Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev.* 20:210–224.

Atkins JF, Wills NM, Loughran G, Wu C-Y, Parsawar K, Ryan MD, Wang CH, Nelson CC. 2007. A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (stop) after addition of glycine and then allowing continued translation (Go). *RNA* 13:803–810.

Balaj L, Lessard R, Dai L, Cho YJ, Pomeroy SL, Breakefield XO, Skog J. 2011. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat Commun.* 2:180.

Brown JD, Ryan MD. 2010. Ribosome "Skipping": "Stop-Carry On" or "StopGo" Translation. In: Atkins JF, Gesteland RF, editors. Recoding: expansion of decoding rules enriches gene expression. New York: Springer. p. 101–122.

de Felipe P, Hughes LE, Ryan MD, Brown JD. 2003. Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide. *J Biol Chem.* 278:11441–11448.

Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439:965–968.

Donnelly MLL, Gani D, Flint M, Monoghan S, Ryan MD. 1997. The cleavage activity of aphtho- and cardiovirus 2A proteins. *J Gen Virol.* 78:13–21.

Donnelly MLL, Hughes LE, Luke G, Li X., Mendoza H, ten Dam E, Gani D, Ryan MD. 2001. The "cleavage" activities of FMDV 2A site-directed mutants and naturally-occurring "2A-Like" sequences. *J Gen Virol.* 82:1027–1041.

Donnelly MLL, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, Ryan MD. 2001. Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal "skip." *J Gen Virol* 82:1013–1025.

Doronina VA, de Felipe P, Wu C, Sharma P, Sachs MS, Ryan MD, Brown J. 2008. Dissection of a co-translational nascent chain separation event. *Biochem Soc Trans.* 36:712–716.

Doronina VA, Wu C, de Felipe P, Sachs MS, Ryan MD, Brown J. 2008. Site-specific release of nascent chains from ribosomes at a sense codon. *Mol Cell Biol.* 28:4227–4239.

Gilbert C, Schaack S, Pace JK 2nd, Brindley PJ, Feschotte C. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464:1347–1350.

Gould PS, Easton AJ. 2005. Coupled translation of the respiratory syncytial virus M2 open reading frames requires upstream sequences. *J Biol Chem.* 280:21972–21980.

Gould PS, Easton AJ. 2007. Coupled translation of the second open reading frame of M2 mRNA is sequence dependent and differs significantly within the subfamily Pneumovirinae. *J Virol.* 81:8488–8496.

Heras SR, Thomas MC, García M, de Felipe P, García-Pérez JL, Ryan MD, López MC. 2006. L1Tc non-LTR retrotransposons from *Trypanosoma cruzi* contain a functional viral-like self-cleaving 2A sequence in frame with the active proteins they encode. *Cell Mol Life Sci.* 63:1449–1460.

Horvath CM, Williams MA, Lamb RA. 1990. Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. *EMBO J.* 9:2639–2647.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 110:462–467.

Kapitonov VV, Tempel S, Jurka J. 2009. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. *Gene* 448:207–213.

Kojima KK, Kapitonov VV, Jurka J. 2011. Recent expansion of a new Ingi-related clade of Vingi non-LTR retrotransposons in hedgehogs. *Mol Biol Evol.* 28:17–20.

Kojima KK, Matsumoto T, Fujiwara H. 2005. Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1. *Mol Cell Biol.* 25:7675–7686.

Luke GA, de Felipe P, Lukashev A, Kallioinen SE, Bruno EA, Ryan MD. 2008. The occurrence, function and evolutionary origins of "2A-like" sequences in virus genomes. *J Gen Virol.* 89:1036–1042.

Luttermann C, Meyers G. 2007. A bipartite sequence motif induces translation reinitiation in feline calicivirus RNA. *J Biol Chem.* 282:7056–7065.

Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol.* 16:793–805.

Meyers G. 2003. Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependent reinitiation mechanism. *J Biol Chem.* 278:34051–34060.

Meyers G. 2007. Characterization of the sequence element directing translation reinitiation in RNA of the calicivirus rabbit hemorrhagic disease virus. *J Virol.* 81:9623–9632.

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr. 1996. High frequency retrotransposition in cultured mammalian cells. *Cell* 87:917–927.

Novikova OS, Blinov AG. 2009. Origin, evolution, and distribution of different groups of non-LTR retrotransposons among eukaryotes. *Genetika* 45:149–159.

Powell ML, Napthine S, Jackson RJ, Brierley I, Brown TD. 2008. Characterization of the termination-reinitiation strategy employed in the expression of influenza B virus BM2 protein. *RNA* 14:2394–2406.

Putnam NH, Butts T, Ferrier DE, et al. (37 co-authors). 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.

Putnam NH, Srivastava M, Hellsten U, et al. (19 co-authors). 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86–94.

Routh A, Domitrovic T, Johnson JE. 2012. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc Natl Acad Sci U S A.* 109:1907–1912.

Ryan MD, Donnelly MLL, Lewis A, Mehrotra AP, Wilkie J, Gani D. 1999. A model for non-stoichiometric, co-translational protein scission in eukaryotic ribosomes. *Bioorganic Chem.* 27:55–79.

Ryan MD, Drew J. 1994. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J.* 134:928–933.

Ryan MD, King AMQ, Thomas GP. 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol.* 72:2727–2732.

Sharma P, Yan F, Doronina V, Escuin-Ordinas H, Ryan MD, Brown J. 2012. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res.* 40: 3143–3151.

Sodergren E, Weinstock GM, Davidson EH, et al. (229 co-authors). 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314:941–952.