




The identification of conserved sequence features of co-translationally decayed mRNAs and upstream open reading frames in angiosperm transcriptomes

Rong Guo¹  | Xiang Yu^{1,2}  | Brian D. Gregory¹ 

¹Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

²School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Correspondence

Brian D. Gregory, Department of Biology, University of Pennsylvania, 433 S. University Ave., Philadelphia, PA 19104.
Email: bdgregor@sas.upenn.edu

Present address

Xiang Yu, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China.

Funding information

National Science Foundation (NSF), Grant/Award Numbers: IOS-2023310, IOS-1849708

Abstract

RNA turnover is essential in maintaining messenger RNA (mRNA) homeostasis during various developmental stages and stress responses. Co-translational mRNA decay (CTRD), a process in which mRNAs are degraded while still associated with translating ribosomes, has recently been discovered to function in yeast and three angiosperm transcriptomes. However, it is still unclear how prevalent CTRD across the plant lineage. Moreover, the sequence features of co-translationally decayed mRNAs have not been well-studied. Here, utilizing a collection of publicly available degradome sequencing datasets for another seven angiosperm transcriptomes, we have confirmed that CTRD is functioning in at least 10 angiosperms and likely throughout the plant lineage. Additionally, we have identified sequence features shared by the co-translationally decayed mRNAs in these species, implying a possible conserved triggering mechanism for this pathway. Given that degradome sequencing datasets can also be used to identify actively translating upstream open reading frames (uORFs), which are quite understudied in plants, we have identified numerous actively translating uORFs in the same 10 angiosperms. These findings reveal that actively translating uORFs are prevalent in plant transcriptomes, some of which are conserved across this lineage. We have also observed conserved sequence features in the regions flanking these uORFs' stop codons that might contribute to ribosome stalling at these sequences. Finally, we discovered that there were very few overlaps between the mRNAs harboring actively translating uORFs and those sorted into the co-translational decay pathway in the majority of the studied angiosperms, suggesting that these two processes might be nearly mutually exclusive in those species. In total, our findings provide the identification of CTRD and actively translating uORFs across a broad collection of plants and provide novel insights into the important sequence features associated with these collections of mRNAs and regulatory elements, respectively.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is Brian D. Gregory (bdgregor@sas.upenn.edu).

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Plant Direct* published by American Society of Plant Biologists and the Society for Experimental Biology and John Wiley & Sons Ltd.

1 | INTRODUCTION

Cells need to constantly adjust mRNA abundance in order to respond properly to different environments or growth stages, which are achieved through a variety of mRNA decay mechanisms. In eukaryotes, mRNAs are believed to first undergo decapping or deadenylation before being degraded from 5' to 3' by exonucleases or from 3' to 5' by the exosome complex (Nagarajan et al., 2013; Schoenberg & Maquat, 2012). Previously, eukaryotic mRNAs undergoing translation were believed to be protected from these decay pathways (Parker, 2012). However, it was first observed in *Saccharomyces cerevisiae* that mRNAs can be decapped and undergo 5' to 3' degradation while still being associated with actively translating ribosomes. This process ensures that the translation of residual ribosomes is not affected by the decay and the final polypeptide is full length and functional (Hu et al., 2009, 2010). Subsequently, through capturing and sequencing degrading intermediates of mRNA molecules bearing the hallmark 5' monophosphate (5P-seq), it was revealed that co-translational 5' to 3' decay is prevalent in *S. cerevisiae* mRNAs and this phenomenon is also conserved in *Saccharomyces pombe* (Pelechano et al., 2015). Specifically, an exonuclease XRN1-dependent three-nucleotide periodicity was observed within the collection of sequencing reads that were captured with a free monophosphate on their 5' end (5'P read ends) within the main open reading frame of many yeast mRNAs. This observation can be explained by XRN1 chasing the last translating ribosome translocating three nucleotides at a time along the length of the protein-coding region (ORF) of mRNAs. Additionally, it was found that 5'P read ends of the co-translationally decayed mRNAs also accumulated at 16 or 17 nucleotides (nt) upstream of the stop codon, suggesting pronounced ribosome pausing at stop codons during translation termination (Pelechano et al., 2015).

In plants, various high-throughput approaches such as genome-wide mapping of uncapped and cleaved transcripts (GMUCT) (Gregory et al., 2008), parallel analysis of RNA ends (PARE) (German et al., 2009), and degradome sequencing (Addo-Quaye et al., 2008) that capture and sequence 5' monophosphate bearing mRNA degradation intermediates have also been developed by a number of different groups. Degradome profiling datasets generated using these approaches provided rich resources for studying co-translational mRNA decay in plants. Based on the three-nucleotide periodicity along the main open reading frame and the 5'P read ends that were found to accumulate at 16 and 17 nt upstream of the stop codons, several groups have revealed that co-translational mRNA decay is prevalent in Arabidopsis, rice, and soybean (Crisp et al., 2017; Hou et al., 2016; Yu et al., 2016). It was also revealed in Arabidopsis that both the exonuclease XRN4 and the cap-binding protein CBP80/ABH1 play important roles in co-translational mRNA decay (Yu et al., 2016). Although co-translational mRNA decay has been proven to be conserved in Arabidopsis, rice, and soybean, it is not clear whether co-translational mRNA decay is also conserved in other plant species or in any various growth stages and environmental conditions. Moreover, the sequence features of the mRNAs undergoing co-translational decay in plant species remain mostly unexplored.

Upstream open reading frames (uORFs) are short open reading frames located in 5' untranslated regions (5' UTRs) of mature mRNAs in eukaryotes. When being translated, they generally repress the translation initiation of downstream main open reading frames (mORFs) via arresting ribosomes (Zhang et al., 2019). Occasionally, mRNA decay is induced because the uORF stop codon can be recognized as a premature stop codon and trigger nonsense-mediated decay (NMD) (Uchiyama-Kadokura et al., 2014). Therefore, uORFs are important *cis* regulatory elements for mORF translation in eukaryotes. In plants, uORFs play important roles in developmental processes as well as various stress responses. For instance, the *HB1* mRNA encodes a homeodomain-leucine zipper transcription factor that participates in hypocotyl elongation in short day conditions in Arabidopsis (Capella et al., 2015). The *HB1* mRNA harbors a uORF, whose translation arrests ribosomes and subsequently represses mORF translation. This translation repression keeps the HB1 protein at low levels under non-short day conditions and avoids aberrant developmental phenotypes that can be caused by high levels of HB1 (Ribone et al., 2017).

However, studies of uORFs in plants have been somewhat limited to Arabidopsis due to a lack of methodologies for identifying these sequence elements in other non-Arabidopsis species. Current uORF identification relies heavily on sequence homology-based methods (Hayden & Jorgensen, 2007; Tran et al., 2008) and ribosome profiling techniques (Ingolia et al., 2012). Although a lot of uORFs in non-Arabidopsis species have been predicted based on sequence homology, other uORFs failed to be identified due to a lack of sequence homology. Moreover, sequence homology-based methods cannot reveal if a predicted uORF can be actively translated. Although ribosome profiling can reveal actively translating uORFs, this technique has only been utilized on a select few plant species (Hsu et al., 2016; Lei et al., 2015; Shamimuzzaman & Vodkin, 2018; Wu et al., 2019). Additionally, a lot of uORF studies have focused on the sequence features of the uORF start codon and how those sequence features are related to uORF translation initiation (Chew et al., 2016; Sachs & Geballe, 2006; Wallace et al., 2020), the sequence features of uORF stop codons, and their relation to ribosome stalling remains mostly unexplored.

In order to study the conservation of CTRD in plants other than Arabidopsis, rice, and soybean, we analyzed degradome datasets of an additional seven angiosperms using our previously established pipeline (Yu et al., 2016). From this analysis, we found that co-translational mRNA decay is not only conserved among the angiosperms but also prevalent among different tissues, developmental stages, and environmental conditions in each species. By analyzing the sequence features of all co-translationally decayed mRNAs in each species, we found that this collection of transcripts displays consistent sequence features across the collection of studied angiosperms, suggesting that they might play important roles in co-translational mRNA decay across the plant kingdom.

Given that degradome datasets can also reveal ribosome dynamics and have been used to identify actively translating uORFs in Arabidopsis (Hou et al., 2016; Yu et al., 2016), they provide an alternative to ribosome profiling for studying translating uORFs in the plants.



Thus, we used this collection of 10 angiosperm degradome datasets to identify actively translating uORFs across a wide array of plant species (Yu et al., 2016). This analysis revealed numerous uORFs that were being actively translated in each set of degradome datasets, indicating that translating uORFs are prevalent in angiosperms as well as in various responses to different environmental conditions and growth stages. Meanwhile, we observed consistent sequence features in the regions flanking these translating uORF's stop codon that might contribute to ribosome stalling in this region. Finally, to better understand the relationship between having actively translating uORFs and being sorted into co-translational mRNA decay, we analyzed the number of mRNAs with actively translating uORFs that are also co-translationally decayed in each dataset. We found that mRNAs that contain actively translating uORFs are rarely co-translational decayed in most of the studied angiosperms, suggesting that actively translating uORFs protect their parent transcripts from this degradation pathway.

2 | RESULTS

2.1 | Co-translational mRNA decay is a conserved process among angiosperms

In order to determine whether CTRD is conserved in plant species other than Arabidopsis, soybean, and rice (Hou et al., 2016; Yu et al., 2016), we analyzed multiple high-quality degradome sequencing datasets of four monocots (*Zea mays*, *Setaria viridis*, *Sorghum bicolor*, and *Brachypodium distachyon*) as well as three dicots (*Solanum lycopersicum*, *Medicago truncatula*, and *Phaseolus vulgaris*) that were obtained from the GEO database (Table 1) (Anderson et al., 2018; Baldrich et al., 2015; Crisp et al., 2017; Devers et al., 2011; Formey et al., 2015; Franke et al., 2018; Jeong et al., 2013; Li et al., 2010; Liu et al., 2014; Shamimuzzaman & Vodkin, 2012; Song et al., 2011; Zhou et al., 2010) using the same pipeline established previously (Yu et al., 2016). We tried to include as many degradome datasets from different conditions, tissues, or growth stages as possible for each species because we also wanted to know if co-translational mRNA decay functions across different conditions and/or tissues within these various species.

: The last four columns indicate the number of total CTRD transcripts identified, actively translating uORFs identified, the percent of CTRD transcripts with actively translating uORFs, and the percent of transcripts with actively translating uORFs sorted into CTRD in the corresponding samples.

Using TSI as the preferred measurement, we identified co-translationally decayed mRNAs in all of the 10 angiosperms, indicating that co-translational mRNA decay is conserved in these plant species and likely most others (Table 1, Figure 1, and Appendix S1). Meanwhile, within the same species, co-translationally decayed mRNAs were identified in every sample regardless of the tissue or conditions of treatment, indicating that co-translational mRNA decay is also conserved across different conditions and might play important roles in different biological processes.

2.2 | Co-translationally decayed mRNAs in different angiosperms display consistent sequence features that are different from those that are not decayed by this pathway

If the sequence features of mRNAs play an important role in determining which mRNAs undergo co-translational decay, the co-translationally decayed mRNAs of different species should display shared sequence features that are different from those of the non-co-translationally decayed mRNAs. In total, we compared the 5' UTR length, CDS length, 5' UTR GC content, CDS GC content, 3' UTR length, 3' UTR GC content, codon usage bias, and amino acid composition bias of the co-translational decayed mRNAs with those of all other mRNAs of each species.

From this analysis, we found that the co-translationally decayed mRNAs have on average significantly (all p -values $< .05$; Wilcoxon test) longer 5' UTRs, CDSs, and 3' UTRs compared to those of the other transcripts not degraded by this mechanism in all the species, except soybean. In soybean, the co-translationally decayed mRNAs have significantly (p -value $< .001$; Wilcoxon test) shorter 5' UTRs, and the CDS length is not significantly different from that of the mRNAs not degraded by this pathway (Figures 2-4). We also found that the 5' UTR GC content of the co-translationally decayed mRNAs is consistently significantly (all p -values $< .001$; Wilcoxon test) higher in every species (Figure 5). However, the CDS GC content of co-translationally decayed mRNAs is significantly (all p -values $< .01$; Wilcoxon test) higher in every dicot but lower in 60% of the monocots, except for *S. bicolor* and *B. distachyon*, whose co-translationally decayed mRNAs have a significantly (p -values $< .01$; Wilcoxon test) higher CDS GC content (Figure 6). Although the co-translationally decayed mRNAs in sorghum, Arabidopsis, and tomato have significantly higher 3' UTR GC content (p -values $< .01$; Wilcoxon test), the 3' UTR GC content of those in the rest of the analyzed angiosperm species is not significantly different compared to mRNAs that are not co-translationally decayed (Figure S1). Thus, there are specific sequence features that define mRNAs degraded by CTRD from those that are not, but sometimes these features vary between the angiosperm species analyzed in this study.

We next compared the codon usage bias in co-translationally decayed mRNAs with that in the other mRNAs of each species. Codon usage bias is defined as the unequal use of synonymous codons during translation (Behura & Severson, 2013). To quantify codon usage bias, we employed the relative synonymous codon usage (RSCU) score. For a specific codon, RSCU is defined as the ratio of the observed frequency to its expected frequency if all the synonymous codons were used equally. These values are not related to the amino acid composition or the abundance ratio of synonymous codons, so it can directly reflect the bias of synonymous codon usage (Sharp & Li, 1986). Given that the start codon, the stop codon, and tryptophan have no synonymous codons, they were excluded from this analysis.

We first calculated the RSCU of each codon in every mRNA. Next, the mRNAs were grouped according to whether they are co-translationally decayed mRNAs or not. The mean RSCU value of every



TABLE 1 Summary of degradome samples analyzed for each species in this study

Species	Accession number	Tissue	Treatment	Sequencing method	Number of CTRD mRNA	Number of actively translating uORFs	Number of mRNAs harboring actively translating uORFs	Number of CTRD mRNAs with actively translating uORFs	% of CTRD mRNAs with actively translating uORFs	% of mRNAs with actively translating uORFs sorted into CTRD
<i>Oryza sativa</i>	GSE17398	3-week-old seedlings	Untreated	PARE	2,654	370	290	37	1.39412208	12.75862069
	GSE19050	Young inflorescence	Untreated	PARE	1210	154	123	13	1.074380165	10.56910569
	GSE66610	15-day-old leaf	Mock treatment 30 min	PARE	1,684	333	265	36	2.137767221	13.58490566
<i>Zea mays</i>	GSE47837	Untreated	Magnaporthe oryzae elicitors 30 min	727	94	84	3	.412654746	3.571428571	
		Mock treatment 120 min	1261	270	236	21	1.665344964	8.898305085		
		Magnaporthe oryzae elicitors 120 min	1043	172	154	17	1.62991371	11.03896104		
		Untreated	658	75	67	1	.151975684	1.492537313		
<i>Brachypodium distachyon</i>	GSE52441	Ear stage I	PARE	674	110	101	2	.296735905	1.98019802	
		Ear stage II	359	44	43	0	0	0		
		Ear stage III	690	58	48	1	.144927536	2.083333333		
		Ear stage IV	777	138	120	11	1.415701416	9.166666667		
<i>Sorghum bicolor</i>	GSE112296	Leaf	PARE	1037	111	92	10	.964320154	10.86956522	
		Stem	996	131	108	5	.502008032	4.62962963		
		Root	1823	549	465	54	2.962150302	11.61290323		
		Panicle	472	165	145	7	1.483050847	4.827586207		
<i>Setaria Viridis</i>	GSE150798	Leaf	PARE	231	124	111	0	0	0	
		Panicle	665	288	240	16	2.406015038	6.666666667		
		Root	1200	361	320	26	2.166666667	8.125		
		Leaf	1401	371	322	16	1.142041399	4.968944099		
<i>Arabidopsis thaliana</i>	GSE47121	Panicle stage 3	PARE	925	126	114	8	.864864865	7.01754386	
		Spikelet stage 3	1220	235	223	18	1.475409836	8.071748879		
		Spikelet stage 4	2138	363	331	60	2.806361085	18.12688822		
Unopened flower buds	GMUCT	2138	363	331	60	2.806361085	18.12688822			

(Continues)



TABLE 1 (Continued)

Species	Accession number	Tissue	Treatment	Sequencing method	Number of CTRD mRNA	Number of actively translating uORFs	Number of mRNAs harboring actively translating uORFs	Number of CTRD mRNAs with actively translating uORFs	% of CTRD mRNAs with actively translating uORFs	% of mRNAs with actively translating uORFs sorted into CTRD
	GSE108852	4-week-old rosette leaves	Control	GMUCT	2187	218	210	24	1.09739369	11.42857143
			Long-term salt stress		1692	236	224	21	1.241134752	9.375
	PRJNA391262	3-week-old whole rosettes	Control	PARE	3268	195	179	57	1.744186047	31.84357542
			Excess light 30 min		1928	125	115	27	1.400414938	23.47826087
			Excess light 60 min		1982	170	157	34	1.715438951	21.65605096
			Excess light 120 min		1367	106	101	20	1.463057791	19.8019802
			Recovery 7.5 min		2330	132	123	31	1.330472103	25.20325203
			Recovery 15 min after 60-min excess light		2687	157	144	27	1.004838109	18.75
			Recovery 30 min after 60-min excess light		2795	159	146	41	1.466905188	28.08219178
			Recovery 60 min after 60-min excess light		3690	216	201	65	1.761517615	32.33830846
<i>Solanum lycopersicum</i>	GSE50085	Leaf of moneymaker strain	Mock treatment	PARE	148	196	178	7	4.72972973	3.93258427
			Tomato yellow leaf curl virus		394	431	390	21	5.329949239	5.384615385
	Leaf of FL505 strain	Mock treatment		434	382	347	19	4.377880184	5.475504323	
		Tomato yellow leaf curl virus		409	404	363	20	4.88997555	5.509641873	

(Continues)



TABLE 1 (Continued)

Species	Accession number	Tissue	Treatment	Sequencing method	Number of CTRD mRNA	Number of actively translating uORFs	Number of mRNAs harboring actively translating uORFs	Number of CTRD mRNAs with actively translating uORFs	% of CTRD mRNAs with actively translating uORFs	% of mRNAs with actively translating uORFs sorted into CTRD
<i>Medicago truncatula</i>	GSE26218	Root	Untreated	PARE	801	131	107	12	1.498127341	11.21495327
	GSE34433	Root	Glomus intraradices	PARE	850	139	113	8	.941176471	7.079646018
<i>Glycine max</i>	GSE25260	Seeds	Untreated	PARE	597	375	347	23	3.852596315	6.628242075
	GSE34433	Seed coat of 25–50 mg fresh weight seeds	Untreated	PARE	296	651	296	8	2.702702703	2.702702703
<i>Phaseolus vulgaris</i>		Seed coat of 100–200 mg fresh weight seeds			67	64	64	0	0	0
		Cotyledon of 50 mg fresh weight seeds			517	247	240	2	.386847195	.833333333
		Cotyledon of 100–200 mg fresh weight seeds			106	77	77	0	0	0
		Cotyledon of 300–400 mg fresh weight seeds			561	776	694	19	3.386809269	2.737752161
<i>Phaseolus vulgaris</i>	GSE67433	Seedlings	Untreated	PARE	2729	529	470	105	3.84756321	22.34042553

Note: The last four columns indicate the number of total CTRD transcripts identified, actively translating uORFs identified, the percent of CTRD transcripts with actively translating uORFs, and the percent of transcripts with actively translating uORFs sorted into CTRD in the corresponding samples.

FIGURE 1 Co-translationally decayed mRNAs are identified in 10 angiosperm species. Sample heatmap of 5'P read ends mapped to flanking 100 nt regions of the stop codon of each transcript for each species. The first nucleotide of stop codon is assigned 0 on the horizontal axis, and each row represents a transcript. The color represents the enrichment of 5'P read ends at each nucleotide relative to the average 5'P read ends of flanking 100 nt region.

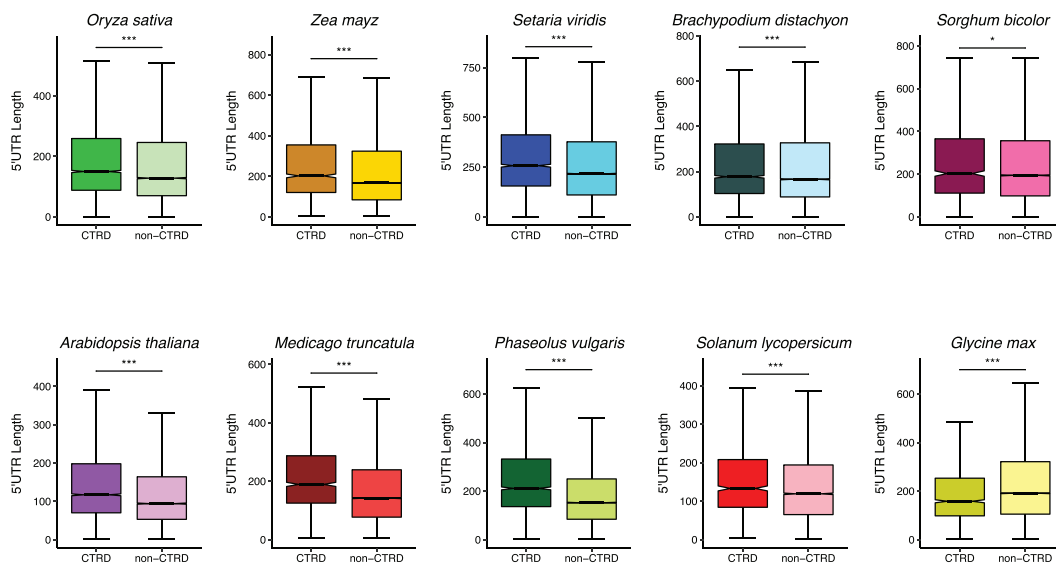
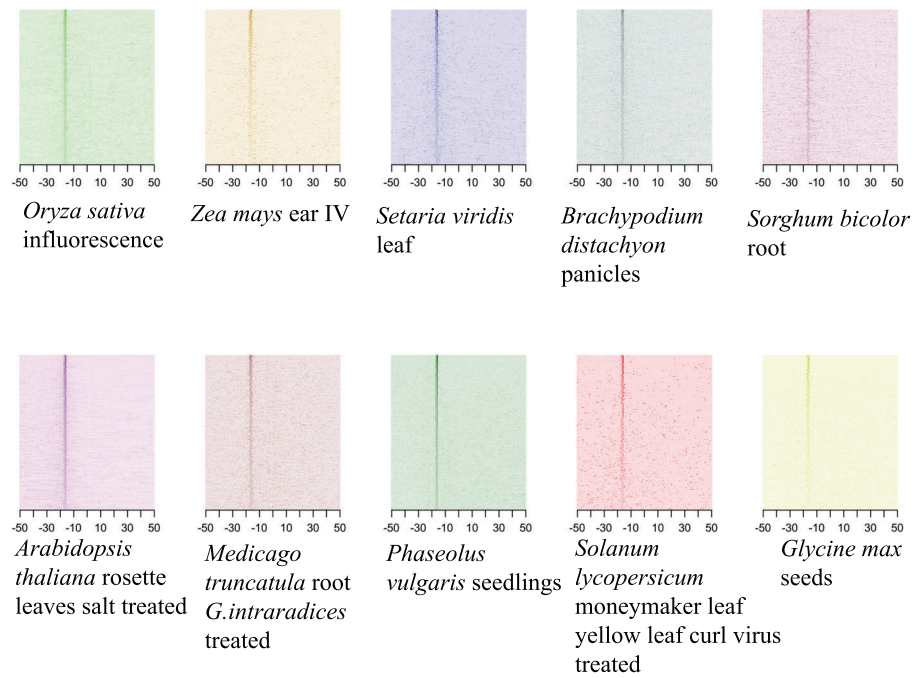


FIGURE 2 CTRD mRNAs have longer 5' UTRs in most of the 10 angiosperm species. *** and * denote p -value ≤ 0.001 and ≤ 0.05 , respectively, as determined by Wilcoxon test.

codon was calculated for both co-translationally decayed mRNAs and those that are not. Then Δ RSCU of every codon was calculated by subtracting the mean RSCU of the non-co-translationally decayed mRNAs from that of the co-translationally decayed mRNAs (Whittle et al., 2011). Therefore, codons with positive Δ RSCU were the ones that were enriched in co-translationally decayed mRNAs compared to its synonymous codons and vice versa. Among the analyzed 59 codons, the CAG (glutamine) and AAG (lysine) codons have positive Δ RSCU values consistently across all the 10 angiosperms, indicating that they are selected for in co-translationally decayed mRNAs compared to their partner synonymous codons CAA (glutamine) and

AAA (lysine) that have negative Δ RSCU values consistently across all the 10 angiosperms (Table 2). Thus, the codons for two positively charged amino acids seem to discriminate mRNAs that undergo CTRD from those that do not.

We finally analyzed the amino acid composition of the peptides encoded by co-translationally decayed mRNAs in each species. By comparing the percentage of each of the 20 amino acids in the co-translationally decayed mRNAs encoded peptides with that of the peptides encoded by those not decayed by this pathway, we found that the co-translationally decayed mRNAs have a significantly (all p -values < 0.01 ; Wilcoxon test) higher percentage of aspartate

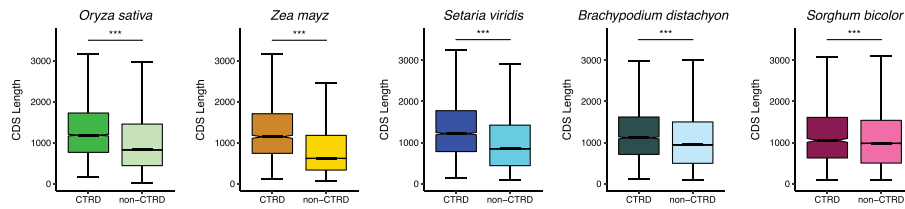


FIGURE 3 CTRD mRNAs have longer CDSs in most of the 10 angiosperm species. *** denotes p -value $\leq .001$, whereas NS denotes p -value $> .05$ as determined by Wilcoxon test.

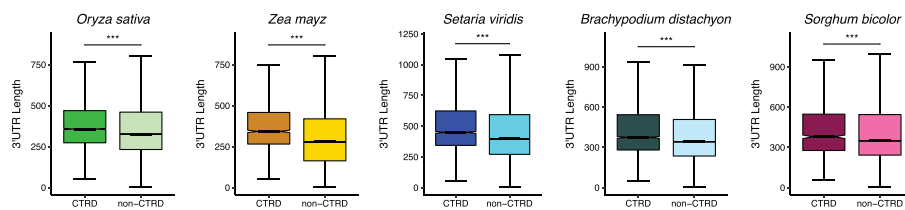
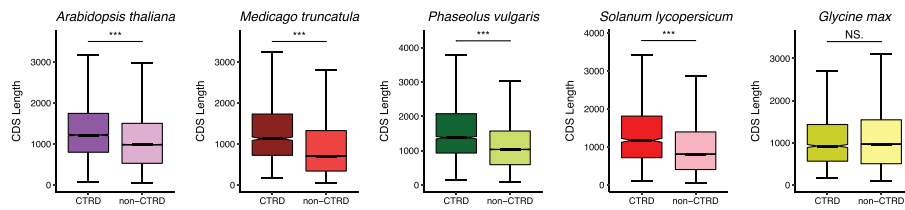


FIGURE 4 CTRD mRNAs have longer 3' UTRs in all 10 angiosperm species. *** denotes p -value $\leq .001$ as determined by Wilcoxon test.

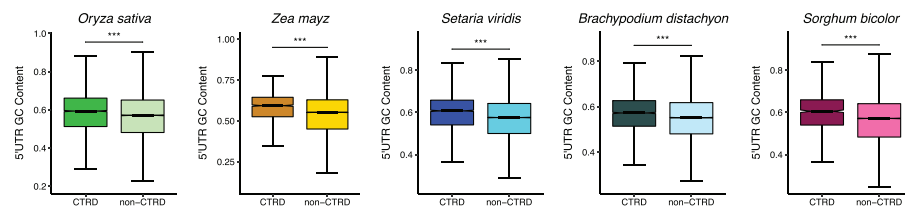
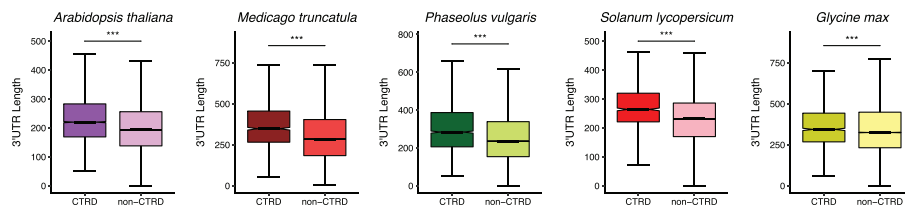
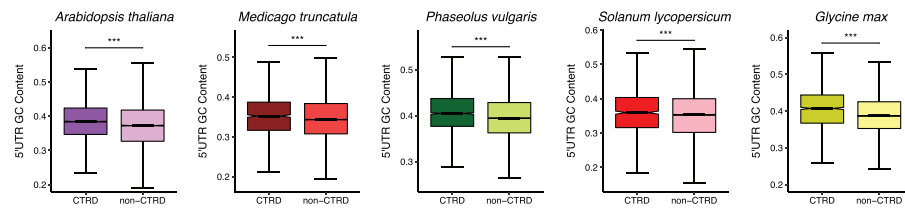


FIGURE 5 CTRD mRNAs have higher 5' UTR GC content in all 10 angiosperm species. *** denotes p -value $\leq .001$ as determined by Wilcoxon test.

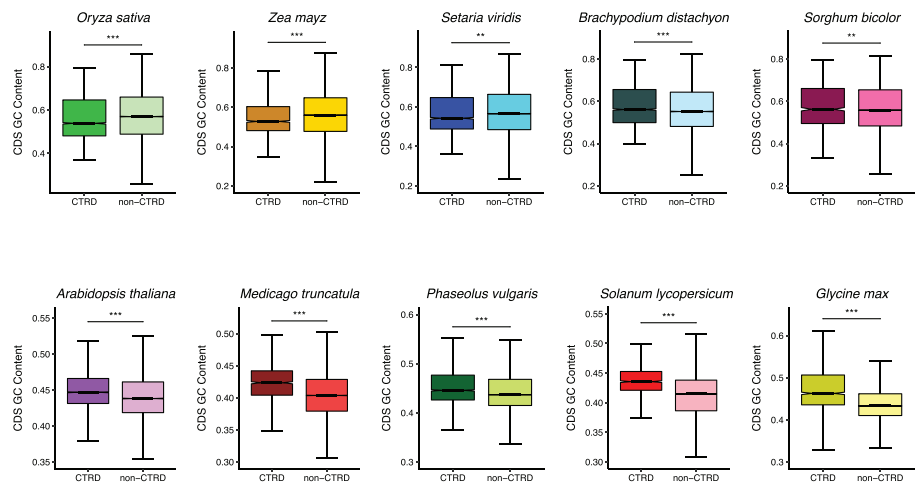


consistently across all the 10 angiosperms (Figure 7). Additionally, the percentage of glutamate is also significantly (all p -values $< .001$; Wilcoxon test) higher in peptides encoded by co-translationally decayed mRNAs in 90% of the interrogated angiosperms, with soybean again being the only outlier for CTRD transcripts not displaying this

sequence feature compared to non-CTRD transcripts (Figure 7). We then wondered if aspartate and glutamate also have positional bias along the mRNAs. Therefore, we compared the distribution of aspartate and glutamate along the peptides encoded by co-translationally decayed mRNAs compared to those not decayed by this mechanism



FIGURE 6 CTRD mRNAs have higher CDS GC content in all dicot species, whereas they have lower CDS GC content in 60% of the monocot species. *** and ** denote p -value ≤ 0.001 and ≤ 0.01 , respectively as determined by Wilcoxon test.



in all 10 species. Surprisingly, for every species, there is no appreciable difference in the distribution of aspartate and glutamate encoded amino acids in the peptides encoded by co-translationally decayed mRNAs compared to those that are not decayed by this pathway (Figure S2). Overall, our findings reveal that the amino acids aspartate and glutamate are highly overrepresented in general in the proteins encoded by transcripts decayed by CTRD compared to those that are not in angiosperms, but their distribution in the peptide is not biased to a particular position. In total, our findings reveal that the mRNAs undergoing co-translational decay in most of the 10 angiosperms display conserved sequence features, suggesting that these sequence features play important roles in determining whether an mRNA undergoes co-translational decay or not in plant transcriptomes.

2.3 | Ribosomes stall longer at some of the codons, and the encoded amino acids enriched in co-translationally decayed mRNAs

After identifying shared sequence features associated with co-translationally decayed mRNAs among the 10 angiosperms, we wanted to know if and how these sequence features impact the translation of these mRNAs. Although it has been reported in multiple plants that the length and the GC content of both the 5' UTR and CDS can impact translation efficiency (Branco-Price et al., 2005; Jiao & Meyerowitz, 2010; Kawaguchi & Bailey-Serres, 2005; Lei et al., 2015; Zhao et al., 2017), we wondered if the synonymous codons and amino acids enriched in co-translationally decayed mRNAs also affect this process. Given that it was reported in yeast and other non-plant eukaryotes that certain nonoptimal codons as well as certain amino acid encoding codons can increase ribosome stalling time during elongation via different mechanisms (Presnyak et al., 2015; Sabi & Tuller, 2015; Yan et al., 2016), we tested if the codons AAG and CAG as well as those encoding aspartate and glutamate that we found were significantly enriched in CTRD transcripts can result in longer ribosome stalling in plant transcriptomes.

Given that ribosome stalling level can be inferred at codon resolution from ribosome profiling datasets (Duncan et al., 2018;

Gerashchenko et al., 2012), we analyzed ribosome stalling at codon AAG, CAG, as well as aspartate and glutamate encoding codons using high-quality ribosome profiling datasets for Arabidopsis shoots, Arabidopsis roots, and tomato roots (Hsu et al., 2016; Wu et al., 2019). To assess ribosome stalling level at the enriched codons, we defined a codon stalling index, where for each codon, we calculated the ratio of the average 5' P read ends of the ribosome footprints when this codon was located at the A and P sites of the ribosome to the median 5' P read ends of all ribosome footprint reads in the flanking 100 nt region. The codon stalling index values of the two codons we found to be enriched in CTRD transcripts were then compared with that of their synonymous codons. We found that the CAG stalling index value is significantly (all p -values < 0.001 ; student t -test) higher compared to that of the synonymous CAA in all the Arabidopsis and tomato ribosome profiling datasets. However, the AAG stalling index is higher compared to that of AAA only in Arabidopsis shoot ribosome profiling dataset but not in those for Arabidopsis or tomato roots (Figure 8a). These results indicated that ribosomes stall longer at CAG (glutamine) compared to its synonymous codon in both Arabidopsis and tomato regardless of the tissue but stall longer at AAG only in Arabidopsis shoots not in the root tissues of these dicot plants.

To assess the ribosome stalling level at the codons encoding the enriched amino acids, we further defined amino acid stalling index to account for the average codon stalling index of all the synonymous codons encoding a specific amino acid. The amino acid stalling index of the enriched amino acids were then compared to the median amino acid stalling index of all the other amino acids. Again, in all the Arabidopsis and tomato samples, we found that aspartate has a higher stalling index compared to the median codon stalling index, whereas glutamate has a lower stalling index when compared to this value for all other amino acids (Figure 8b), indicating that ribosomes stall longer only when decoding aspartate in both Arabidopsis and tomato transcriptomes. Taken together, we have revealed that the ribosomes stall longer at both of the synonymous codons enriched in co-translationally decayed mRNAs in at least one tissue for which translation has been assessed in plant species, whereas they only stall longer when decoding aspartate, which is one of the enriched amino acids in co-translationally decayed mRNAs. Overall, these findings suggest

TABLE 2 Difference in the codon usage in co-translationally decayed mRNAs for the 10 angiosperm species analyzed in this study

Codons	Amino acid	Δ RSCU of each species									
		<i>O. sativa</i>	<i>Z. mays</i>	<i>S. bicolor</i>	<i>B. distachyon</i>	<i>S. viridis</i>	<i>A. thaliana</i>	<i>G. max</i>	<i>M. truncatula</i>	<i>P. vulgaris</i>	<i>S. lycopersicum</i>
CAA	Q	-.0689031	-.0565878	-.1082412	-.112391354	-.0358012	-.09572851	-.1768285	-.1034408	-.123697345	-.1590265
CAG	Q	.10756287	.10957125	.11698083	.132249614	.0776722	.133718189	.17392966	.14548041	.13931587	.215322
AAA	K	-.019631	-.035043	-.0826303	-.08855451	-.0135863	-.02537205	-.1421933	-.0739033	-.042429003	-.1115778
AAG	K	.06026598	.11140606	.09428091	.120063158	.06655169	.040456902	.14736714	.08595996	.048861817	.13523297

that features of CT RD mRNAs that distinguish them from other transcripts affect the process of translation, which is likely a feature that defines this class of mRNAs for this specific degradation pathway in plants.

2.4 | Actively translating uORFs are prevalent among angiosperm species, and some are conserved across this lineage

It is notable that actively translating uORFs are rarely studied in non-*Arabidopsis* plant species due to a lack of ribosome profiling data available for those plants, but we and others have found that degradome datasets can also be used to identify actively translating uORFs in plant transcriptomes (Hou et al., 2016; Yu et al., 2016). Thus, we took full advantage of the collection of degradome datasets for the 10 angiosperm species and identified numerous actively translating uORFs using a previously reported method (Yu et al., 2016). Briefly, putative uORFs were first predicted from 5' UTR sequences of each species. Given that it was reported that the near-consensus codons CTG and ACG might also initiate uORF translation in plants (Laing et al., 2015; Liu et al., 2013; Yu et al., 2016), we also included the putative uORFs with these near-consensus start codons in our analysis. We then calculated the TSI for each predicted uORF and considered the uORFs with TSI greater than three to be actively translating uORFs. Using this method, we identified hundreds of actively translated uORFs in every plant species studied (Table 1), indicating that actively translated uORFs are prevalent in angiosperm transcriptomes as well as among different developmental and environmental conditions that plants experience. To understand how conserved these actively translating uORFs are among the angiosperms, we paired the mRNAs harboring the actively translating uORFs in each species with their *Arabidopsis* ortholog. For each annotated *Arabidopsis* ortholog, we then counted the number of species where an actively translating uORF was identified. In total, we found 13 mRNAs that contain actively translating uORFs in over five angiosperm species (Table 3). Among these 13 mRNAs, the uORFs of *bZIP53*, *bZIP44*, putative *eIF5*, *CIPK6*, and *NIP5* mRNAs were reported to be actively translated only in *Arabidopsis* to date (Hou et al., 2016; Tanaka et al., 2016; Weltmeier et al., 2009). However, our findings indicated that the translation of these uORFs is also prevalent in other angiosperms, with the *bZIP53* uORF demonstrating active translation in all of the 10 interrogated angiosperms (Figure 9a).

We next studied if the peptides encoded by the uORFs of these 13 mRNAs are conserved in each species. Given that the shortest peptide derived from translation in plants is 10 amino acids in length (De Coninck et al., 2013; Hsu & Benfey, 2018), we excluded the uORFs that are shorter than 33 nt from this analysis. According to alignment results, peptides encoded by *bZIP53*, *bZIP44*, and *CIPK6* uORFs showed very high levels of conservation at their C-terminus, whereas their N-terminus appeared divergent (Appendix S2A top three panels). We also noticed that the start codon of *bZIP53* and *bZIP44* uORFs in many of the species were near-consensus start

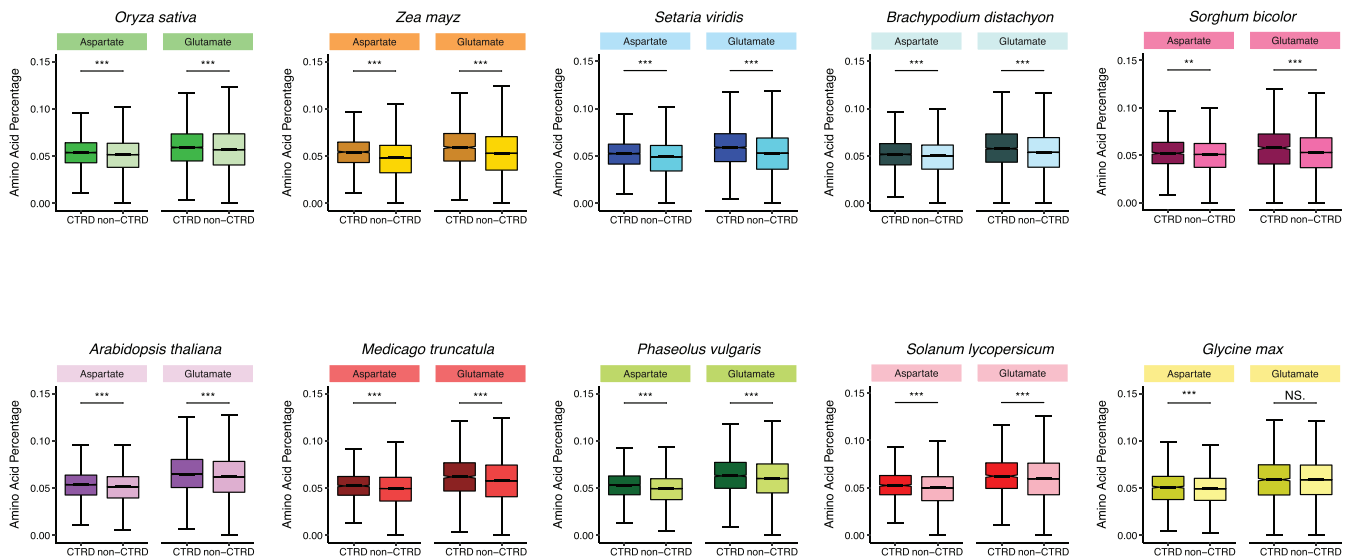


FIGURE 7 Peptides encoded by CTRD mRNAs have a higher percentage of aspartate as well as glutamate in most of the 10 angiosperm species. *** and ** denote p -value $\leq .001$ and $\leq .01$, respectively, whereas NS denotes p -value $> .05$ as determined by Wilcoxon test.

codon sequences, whereas the majority of *CIPK6* uORFs start codons are the consensus ATG sequence. Apart from the peptides derived from those three uORFs, we did not observe high level of conservation in the peptides encoded by uORFs of the other 10 mRNAs (Appendix S2A).

Although the majority of these 13 mRNAs contain translating non-peptide encoding uORFs in only one or two species (Additional Dataset S3), we noticed that *bZIP53*, *SIP1*, and *NIP5* mRNAs contain this type uORFs in five species, suggesting that the non-peptide encoding uORFs of these three mRNAs might serve important functions. Therefore, we wanted to know if their nucleotide sequences are conserved. By aligning these short uORFs along with their 15 nt flanking regions relative to their start and stop codon, we found that *bZIP53* and *SIP1* uORFs are only slightly conserved in the regions flanking their start codon (Appendix S2B). However, *NIP5* uORFs and their entire flanking regions are highly conserved among all the seven angiosperms in which it is actively translated (Figure 9b,c). Taken together, our analysis further indicated that degradome datasets are valuable resources for identifying actively translating uORFs when ribosome profiling data is not available for a specific organism. We have also confirmed that actively translating uORFs are prevalent in angiosperms and might potentially have important physiological functions, but this hypothesis will require further inquiry.

2.5 | The regions flanking the stop codons of actively translating uORFs display consistent sequence features

Given that these identified actively translating uORFs are all able to stall ribosomes sufficiently to measure their accumulation at these stop codons, we wondered if there are any sequence features shared

by the flanking regions of these uORF stop codons that might contribute to ribosome stalling. We first characterized the secondary structure of regions around the stop codons of the translating uORFs in each species via calculating the ensemble free energy (EFE). Specifically, the EFE was calculated using ViennaRNA Package in 35 nt sliding windows for the 100 nt regions flanking the stop codon of the identified uORFs. A lower EFE value means a more stable secondary structure (Chew et al., 2016; Gruber et al., 2015). For all the monocot and dicot species, we found a sudden slight drop of EFE values around the uORF stop codons, indicating the presence of a more stable secondary structure compared to the other regions. Although all the monocots seem to have lower EFE value compared to dicots, the drop of EFE values around their uORF stop codons are consistent in these plant species as well (Figure 9d). Thus, a drop in secondary structure following the stop codon seems to be a hallmark of actively translating uORFs across the plant lineage.

Next, we examined the sequence context in the 10 nt upstream regions of the uORF stop codons. From this analysis, we observed that adenosine is underrepresented at the -1 position relative to the first nucleotide of uORF stop codon consistently in all of the angiosperms (Figure 10, boxed nt 10), whereas guanosine and/or cytosine is overrepresented also at this -1 position consistently in all of the monocot species (Figure 10). In all the dicots, thymine is overrepresented at both -1 and -2 positions relative to the first nucleotide of uORF stop codons (Figure 10, boxed nt 9 and 10). We also noticed that thymine is overrepresented, whereas guanosine was underrepresented at -8 position relative to the first nucleotide of uORF stop codons consistently in all the dicot species (Figure 10, boxed nt 3). Taken together, we revealed that the regions flanking the translating uORFs' stop codons are consistently more structured in all the angiosperm species interrogated in this study. Additionally, although sequence patterns specific to the flanking regions of translating

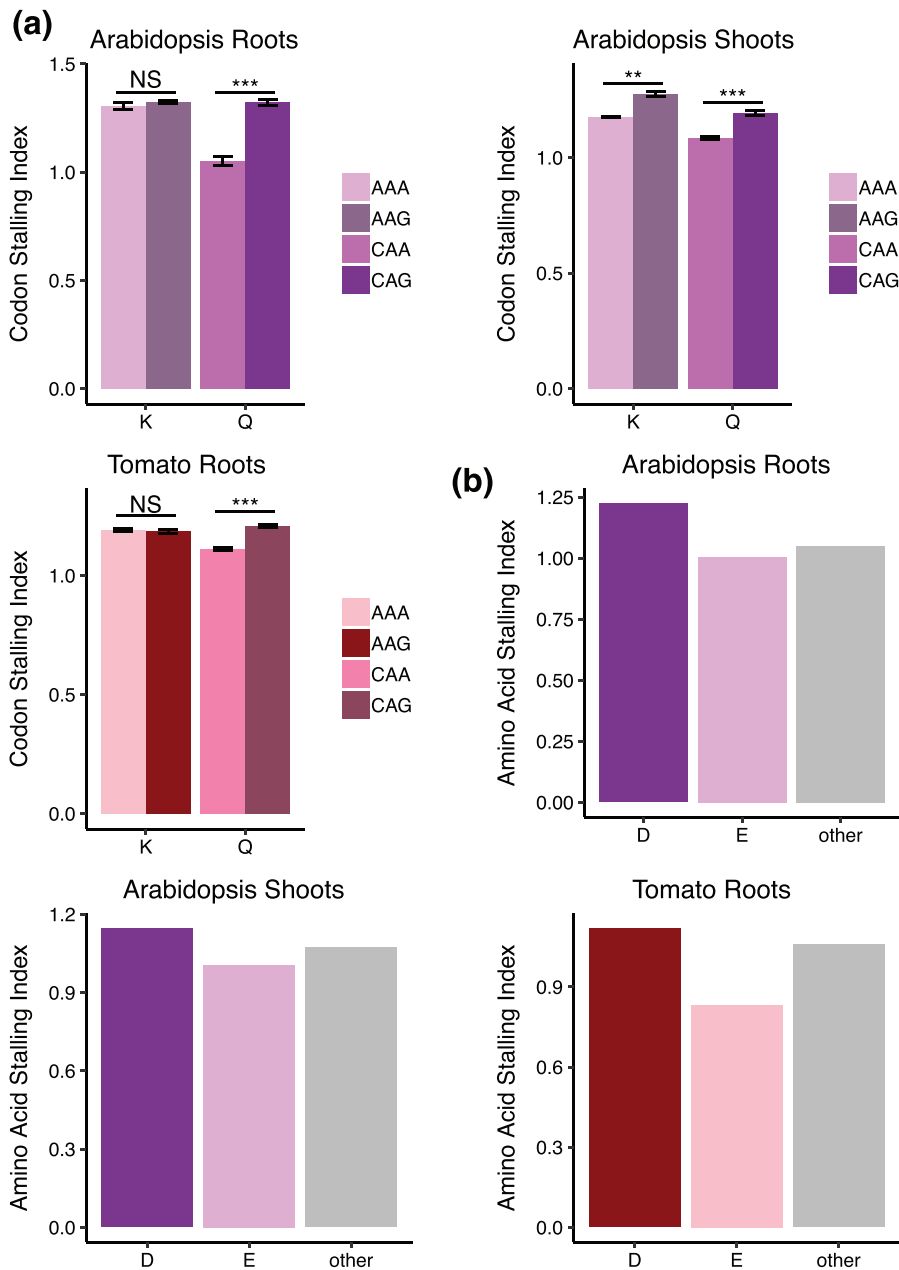


FIGURE 8 Ribosomes pause longer at some of the enriched codons and amino acids identified as overrepresented in co-translationally decayed mRNAs. (a) Codon stalling index for synonymous codon pairs CAG/CAA and AAG/AAA calculated using ribosome profiling data for Arabidopsis roots, Arabidopsis shoots, and tomato root. (b) Amino acid stalling index for aspartate and glutamate calculated using ribosome profiling data for Arabidopsis root, Arabidopsis shoots, and tomato roots. Median represents the median of other amino acid stalling indexes calculated using the specified sample. *** and ** denotes p -value $\leq .001$ and $.01$, respectively, whereas NS denotes p -value $> .05$ as determined by Student's t -test. Error bars represent the standard error for the mean of three biological replicates.

uORFs' stop codons are observed, the patterns are divergent between monocots and dicots yet conserved within the monocot and dicot species interrogated by this study.

2.6 | Translating uORFs and co-translational mRNA decay mostly mutually exclusive in the majority of the studied angiosperms

Finally, given that translating uORFs are able to inhibit downstream mORF translation in eukaryotes (Zhang et al., 2019), we wondered whether and how the actively translating uORFs impact downstream co-translational mRNA decay. To answer this question, we overlapped the mRNAs harboring translating uORFs with those identified to be co-translationally decayed in all of the analyzed

datasets. We found that less than 5% of co-translationally decayed mRNAs harbored actively translating uORFs in all of the 10 angiosperms and across all of the conditions, suggesting that the majority of the CTRD mRNAs are devoid of translating uORFs (Figure 11 and Table 1). Although only about 10% of mRNAs harboring actively translating uORFs are also sorted into the co-translational decay pathway in eight out of the 10 angiosperms, this percentage increased to over 20% in one Arabidopsis dataset (PRJNA391262) and the common bean dataset (Figure 11 and Table 1), where a lot more CTRD mRNAs and lesser actively translating uORFs were identified compared to the other datasets. Together, these results indicate that mRNAs containing translating uORFs are rarely degraded co-translationally in the majority of studied plant transcriptomes in an abundance of varying conditions, with a few minor exceptions.

TABLE 3 Conserved actively translating uORFs identified in degradome datasets for 10 angiosperm species

Arabidopsis homologs	<i>A. thaliana</i> _uORF	<i>O. sativa</i> _uORF	<i>Z. mays</i> _uORF	<i>G. max</i> _uORF	<i>M. truncatula</i> _uORF	<i>P. vulgaris</i> _uORF
AT3G62420	1	1	1	1	1	1
AT1G36730	1	1	1	1	1	0
AT4G02080	1	1	0	1	1	1
AT1G75390	1	1	1	0	1	1
AT4G10380	0	1	1	0	1	1
AT4G30960	1	1	0	1	1	1
AT1G09640	1	1	0	1	0	1
AT4G31170	1	1	0	1	0	0
AT1G65840	1	0	0	1	0	0
AT5G64170	1	1	0	0	1	0
AT4G40060	1	1	0	1	1	1
AT5G11500	1	1	1	0	0	0
AT5G58380	0	1	1	1	1	0

In the table, a "1" value indicates, the uORF from the Arabidopsis ortholog is identified as actively translating in the species, whereas "0" indicates, the uORF is not identified as actively translating. Sum indicates the total number of species that the uORF is identified to be actively translating within.

TABLE 3 (Continued)

Arabidopsis homologs	<i>S. lycopersicum</i> _uORF	<i>B. distachyon</i> _uORF	<i>S. bicolor</i> _uORF	<i>S. viridis</i> _uORF	Sum	Gene_Name
AT3G62420	1	1	1	1	10	bZIP53
AT1G36730	1	0	1	1	8	Putative eIF5
AT4G02080	1	0	1	0	7	SAR1c
AT1G75390	1	0	1	0	7	bZIP44
AT4G10380	1	0	1	1	7	NIP5
AT4G30960	0	0	1	0	6	CIPK6
AT1G09640	0	1	0	1	6	EF1B
AT4G31170	1	0	1	1	6	RAF28
AT1G65840	1	1	1	1	6	PAO4
AT5G64170	1	0	1	1	6	LNK1
AT4G40060	1	0	0	0	6	HB16
AT5G11500	1	0	1	1	6	Coiled-coil protein
AT5G58380	0	0	1	1	6	SIP1

In the table, a "1" value indicates, the uORF from the Arabidopsis ortholog is identified as actively translating in the species, whereas "0" indicates, the uORF is not identified as actively translating. Sum indicates the total number of species that the uORF is identified to be actively translating within.

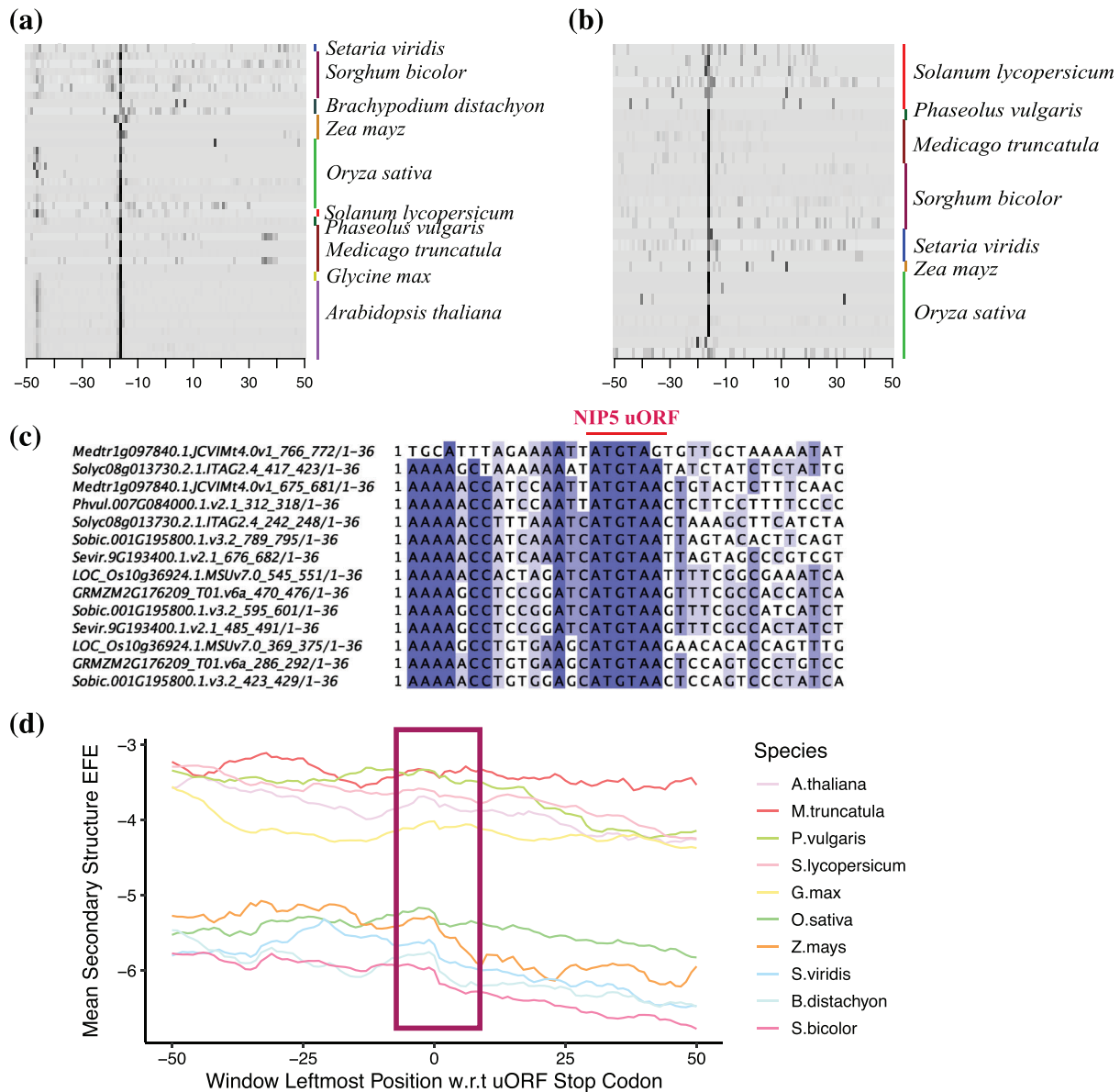


FIGURE 9 Actively translating uORFs are identified in degradome datasets for 10 angiosperm species. Heatmap of 5'P read ends mapped in 100 nt flanking regions of the stop codon of (a) *bZIP53* uORF and (b) *NIP5* uORF in each of the 10 studied angiosperm species. The first nucleotide of uORF stop codon is assigned 0 on the horizontal axis, and each row represents the uORF in a specific sample for each species (specified on the right of each figure), which corresponds to a specific developmental or treatment condition (Table 1). The color represents the enrichment of 5'P read ends at each nucleotide relative to the average 5'P read ends of the flanking 100 nt region. (c) Alignment of *NIP5* uORF nucleotide sequences along with its flanking region for the eight angiosperm species in which is found to be actively translated. The asterisk indicates a conserved nucleotide at those specific nucleotide positions. (d) Mean secondary structure entropy free energy (EFE) of 100 nt flanking regions of actively translating uORF stop codons in all of the 10 angiosperm species. The boxed region denotes the regions that demonstrate a decrease in mean EFE near the uORF stop codons.

3 | DISCUSSION

After the discovery of co-translational mRNA decay in yeast (Hu et al., 2009, 2010; Pelechano et al., 2015), several studies indicated CTRD was also prevalent in the transcriptomes of the plants *Arabidopsis*, rice, and soybean (Crisp et al., 2017; Hou et al., 2016; Yu et al., 2016). However, it was unclear if CTRD is conserved in any other plant species. In this study, we provided evidence that this

decay pathway is also conserved in seven other angiosperms and throughout different tissues, growth stages, and treatment conditions (Figure 1, Table 1 and Appendix S1). Although our results suggest that CTRD is widely conserved across angiosperms, it is still unknown whether this pathway is conserved among other subgroups within the plant kingdom. Therefore, more degradome datasets of plant species from different subgroups will be needed in order to answer this question. Interestingly, CTRD has also been found to exist in human cells

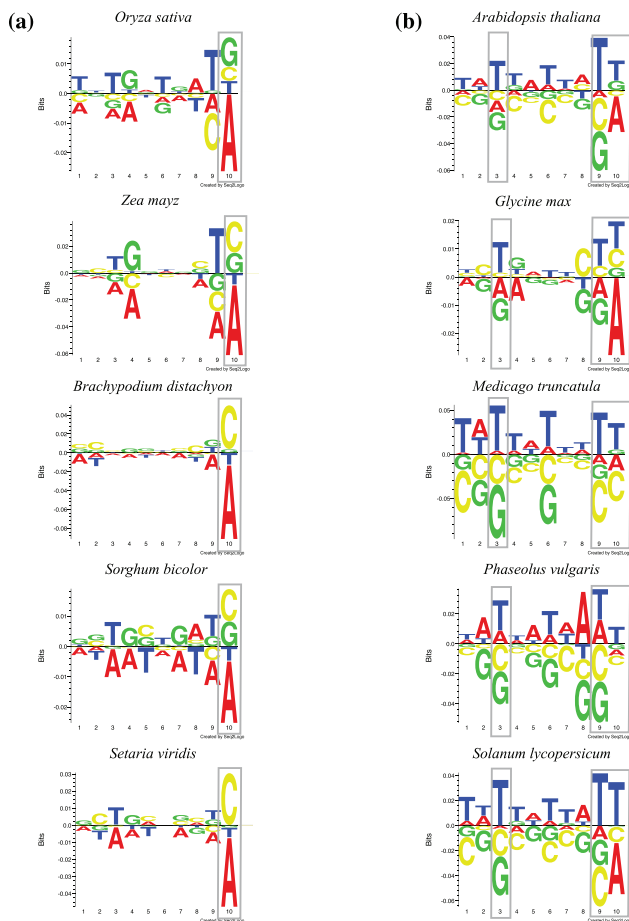


FIGURE 10 Consistent nucleotide patterns are observed in the 10 nt region upstream of actively translating uORF stop codons. Nucleotides shown below the horizontal axis represent the repressed nucleotides at a certain location, whereas those above are overrepresented. Position 10 indicates the nucleotide right before the “T” in the stop codon. Boxed areas indicate the locations where consistent patterns were observed.

(Ibrahim et al., 2018), suggesting that this degradation mechanism is likely conserved among all eukaryotes.

Aiming at understanding the factors that determine which mRNAs undergo co-translational decay, we looked for sequence features that are shared by the co-translationally decayed mRNAs of each angiosperm species. We found no consistent patterns of 3' UTR GC content for CTRD mRNAs compared to non-CTRD mRNAs (Figure S1), suggesting that 3' UTR GC content might not play a role in determining which mRNAs are sorted into this degradation pathway. However, we found several shared sequence features among CTRD mRNAs in these 10 angiosperms. Specifically, we found that the CTRD mRNAs generally have longer 5' UTRs and CDSs consistently in nine of the 10 angiosperms, with the exception being CTRD in soybean transcriptomes (Figures 2–3). Additionally, the 3' UTRs of CTRD mRNAs are also consistently longer in all 10 angiosperms (Figure 4). Furthermore, the 5' UTR GC content of co-translationally decayed mRNAs is consistently higher in all of the 10 angiosperms,

whereas the CDS GC content is higher in the CTRD mRNAs of dicots but lower in 60% of the monocots, with *S. bicolor* and *B. distachyon* being exceptions (Figures 5–6). Several studies have indicated that these shared sequence features can affect mRNA translation efficiency. Thus, these results suggest that variations in translation may be a defining feature of CTRD mRNAs.

In fact, it was previously reported in several angiosperms that mRNA 5' UTR length, CDS length and 5' UTR GC content are all negatively correlated with translation efficiency (Branco-Price et al., 2005; Jiao & Meyerowitz, 2010; Kawaguchi & Bailey-Serres, 2005; Lei et al., 2015; Zhao et al., 2017). Meanwhile, mRNAs with higher CDS GC content have higher translation efficiency in both rice and maize (Lei et al., 2015; Zhao et al., 2017), implying that CDS GC content is positively correlated with translation efficiency in monocots. However, in Arabidopsis, the translation efficiency does not show strong correlation with CDS GC content (Zhao et al., 2017), suggesting that CDS GC content might not affect translation efficiency in dicot species. Although it was found in mammalian cells that longer 3' UTRs also correlate with lower translation efficiency (Tanguay & Gallie, 1996), it is still unclear how 3' UTR length affects translation efficiency in plants in general. Therefore, given that the co-translationally decayed mRNAs in most of these 10 angiosperms have three sequence features (longer 5' UTRs, longer CDSs, and higher 5' UTR GC content) that are associated with low translation efficiency, it can be inferred that these co-translationally decayed mRNAs are very likely to have lower translation efficiency in angiosperms.

By comparing the RSCU value of the codons in the co-translationally decayed mRNAs with that of their synonymous codons, we found that codons CAG and AAG are enriched in CTRD mRNAs relative to their synonymous codons consistently in all the 10 angiosperms (Table 2). By comparing the percentage of each amino acid in the peptides encoded by CTRD mRNAs with that in the peptides encoded by all other transcripts, we found that aspartate was enriched in peptides encoded by co-translationally decayed mRNAs in every angiosperm, whereas glutamate is enriched in nine of the 10 the angiosperms, with the exception again being soybean (Figure 7). Also, we did not observe any difference in the distribution of aspartate and glutamate in peptides encoded by co-translationally decayed mRNAs and those encoded by all other transcripts (Figure S2). Although we found individual amino acids that are enriched in peptides encoded by co-translationally decayed mRNAs, we did not find any enriched amino acid motifs. Interestingly, it was discovered in yeast that serine was enriched at the position immediately after methionine in peptides encoded by CTRD mRNAs (Pelechano et al., 2015). These results imply that specific amino acids play roles in determining which mRNAs undergo co-translational decay in yeast and angiosperm plant species, but it is not the same amino acids in these differing eukaryotic organisms.

By comparing ribosome stalling levels based on Arabidopsis and tomato ribosome profiling datasets corresponding to the specific codons and amino acids enriched in CTRD mRNAs compared to all others, we found that ribosomes pause longer at the enriched codon CAG compared to its synonymous codon in both Arabidopsis and

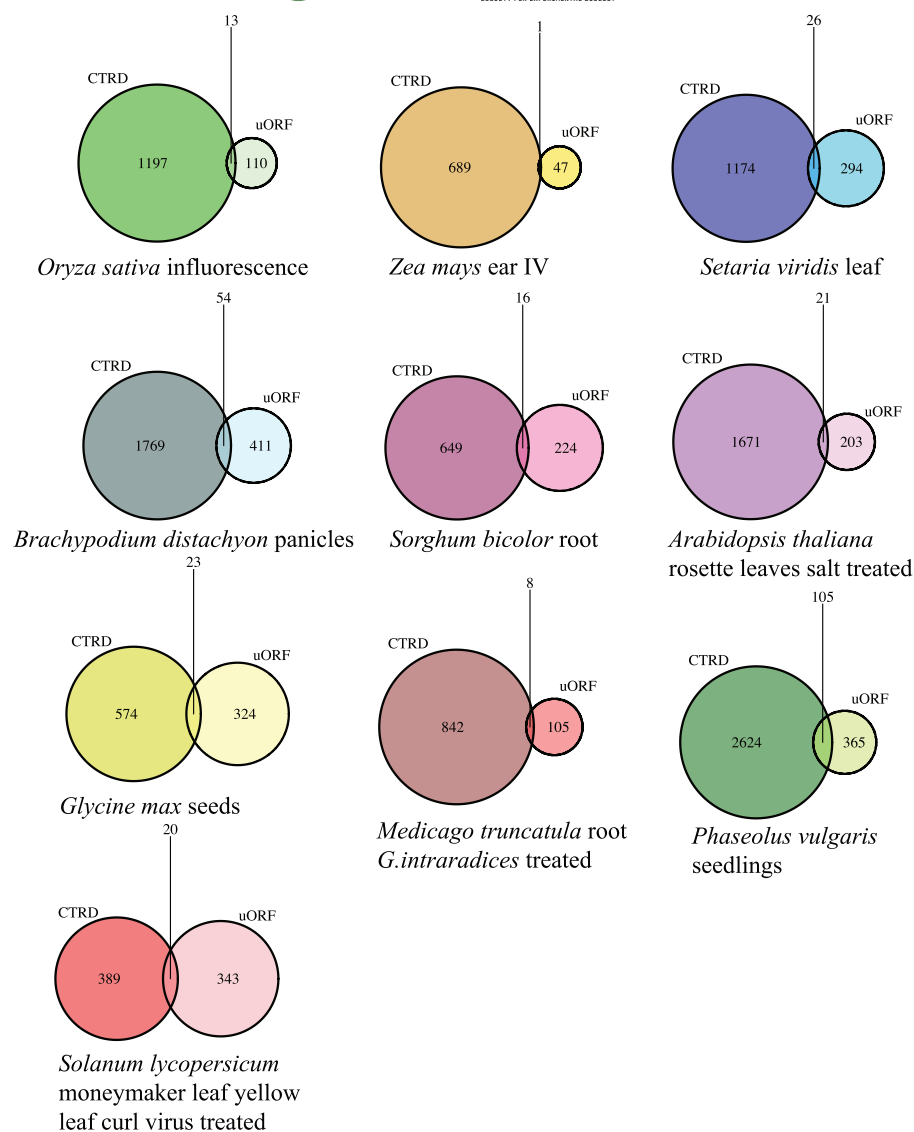


FIGURE 11 uORF translation and co-translational mRNA decay are mostly mutually exclusive in the transcriptomes of most of the 10 studied angiosperms. “uORF” indicates mRNAs with actively translating uORFs, whereas “CTRND” indicates co-translationally decayed mRNAs.

tomato. However, it was surprising to find that ribosomes pause longer at AAG only in *Arabidopsis* shoots as compared to *Arabidopsis* and tomato root ribosome profiles (Figure 8a). These results suggest that the ribosome stalling level could be dependent on the tissue type in which CTRD is being profiled, and thus depending on the tissue transcripts containing more CAG and/or AAG in their primary sequence is likely to decrease the overall ribosome elongation rate on the corresponding mRNAs and initiate CTRD of those transcripts. This hypothesis is worth testing with future experiments. Compared with the ribosome stalling level at the enriched codons in CTRD mRNAs, ribosome stalling levels at the enriched amino acids aspartate and glutamate are very consistent among *Arabidopsis* and tomato. In fact, in both *Arabidopsis* and tomato, the ribosome stalling level is higher only at aspartate codons but not those encoding glutamate (Figure 8b). Interestingly, the ribosome stalling patterns for codons and amino acids observed in *Arabidopsis* and tomato are consistent with what was observed in other eukaryotes. For instance, in yeast, codons AAG and CAA were considered optimal codons, which had a faster

ribosome translocation rate, whereas AAA and CAG were nonoptimal codons with a slower translocation rate (Presnyak et al., 2015). It was also observed in *Drosophila*, human, and yeast cells that ribosomes stall longer when encountering the codons for aspartate in mRNA sequences (Sabi & Tuller, 2015). These results suggested that the observed ribosome stalling patterns in *Arabidopsis* and tomato might be conserved across eukaryotes, including monocots. These observations lead to the inference that ribosome elongation is slower on co-translationally decayed mRNAs, because these mRNAs harbor more codons that could either stall ribosomes longer by themselves or encode amino acids that result in increased ribosomes stalling.

Since it was reported in *Arabidopsis* that XRN4 selectively degraded mRNAs that contain particular sequence motifs (Rymarquis et al., 2011), we wanted to know if there are any overrepresented motifs in the co-translationally decayed mRNAs of each species since XRN4 is the exonuclease that is known to function in the co-translational decay process (Yu et al., 2016). We both searched for the motifs that were found to be enriched in XRN4 targets according to

the previous study (Rymarquis et al., 2011) and conducted de novo motif searches in the CTRD mRNAs of each species. However, none of the searches yielded enriched motifs relative to all other mRNAs in the 10 plant species, indicating that XRN4 selectivity does not play a role in determining co-translational decay targets. Thus, it seems to be the other features identified by this study that truly delineate CTRD mRNAs from those that are not degraded by this pathway.

It is worth noting that the co-translationally decayed mRNAs of soybean tended to be the only outlier group when identifying sequence features that were specific to CTRD mRNAs in the 10 angiosperm species interrogated for this study. Given that the soybean degradome datasets that were analyzed in this study were the only ones generated from seeds (embryonic tissues), the observed sequence feature difference in CTRD targets could imply a massive difference between embryonic tissue and other somatic tissues in the regulation of this specialized degradation mechanism. However, this could also be interpreted as a species-specific difference between soybeans and the other nine angiosperms that were analyzed. Overall, the major differences observed for soybeans, whether tissue or species specific, need to be further interrogated in future studies.

Despite these observed outlier results in soybean, most co-translationally decayed mRNAs in angiosperms have conserved sequence features, suggesting a conserved initiation mechanism for this degradation pathway in angiosperms, and likely other plant species. Given that the majority of these conserved sequence features are associated with low translation efficiency and/or slower ribosome elongation, we hypothesize that CTRD is induced on the portion of mRNAs in plant transcriptomes that display very low translation activity (Figure 12a). It will be interesting to investigate how exactly does low translation efficiency result in CTRD within the plant transcriptome in the future experiments.

In addition to co-translationally decayed mRNAs, we also identified numerous previously undiscovered actively translating uORFs using degradome datasets of many non-Arabidopsis angiosperms. These results indicate that actively translating uORFs are prevalent among angiosperm transcriptomes and can be identified in these plant species within different tissues, growth stages, and environmental conditions (Table 1). Intriguingly, we were able to confirm that many of the uORFs that were only predicted previously using sequence homology-based methods are being actively translated in various angiosperm transcriptomes where they were not previously known to be functional. For instance, we confirmed that *bZIP53*, *bZIP44*, and *CIPK6* uORFs, which have been speculated to be present and active in several plant species (Takahashi et al., 2012; Vaughn et al., 2012; Weltmeier et al., 2009; Wiese et al., 2004), are actively translated in more angiosperms than previously predicted, where they are actively translated in 10, 7, and 6 angiosperms, respectively (Figure 9a and Table 3). These results suggest that numerous conserved uORFs have important regulatory functions in various biological processes across the plant lineage. This hypothesis will need further testing in the future to determine the biological consequences of these potential regulatory sequences.

It is notable that sequence alignment of the peptides encoded by these uORFs of various angiosperms indicated that the resultant peptide in each species is only conserved in its C-terminal region (Appendix S2A first three panels), indicating that any functionality of these small peptides is likely linked to these C-terminal regions. It was previously suggested that ribosomes could be stalled at uORF stop codons via the interaction between the peptide being translated using the uORF as a template and the exit tunnel of the ribosome (Bhushan et al., 2010). Therefore, it might be possible that the conserved C-terminus of these uORF peptides in each species encodes an amino

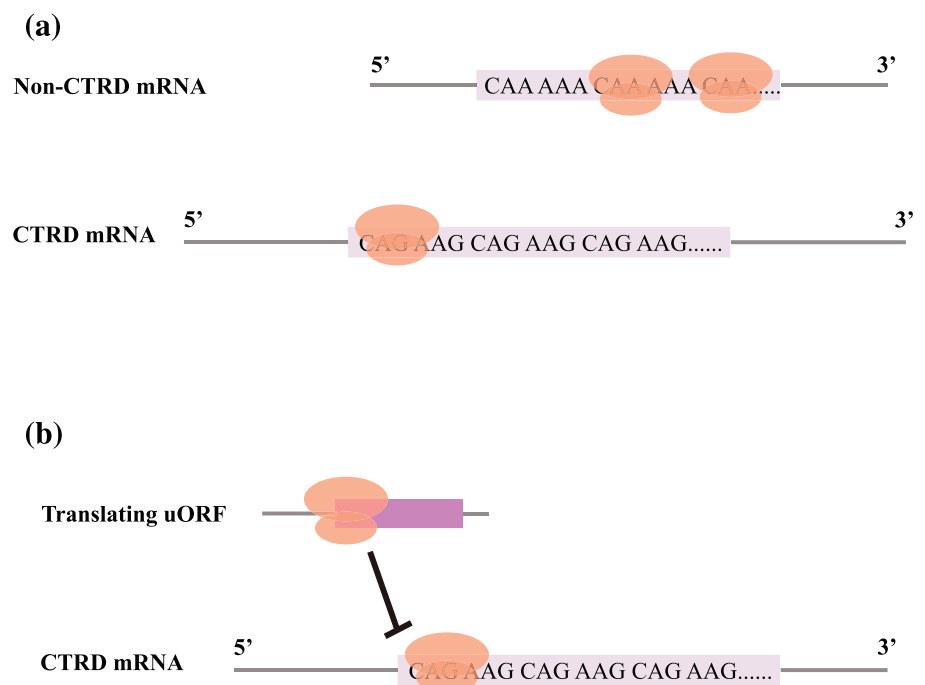


FIGURE 12 Models for co-translational mRNA decay and uORF translation in angiosperms.

(a) Hypothesized initiation mechanism for co-translational mRNA decay. mRNAs with longer 5' UTRs, longer CDSs, as well as enriched with codons CAG and AAG are more likely to have lower translation efficiency or slower elongation during translation, which triggers co-translational mRNA decay. (b) In most of the angiosperms, translating uORFs are likely to inhibit downstream translation, thereby inhibiting co-translational mRNA decay.

acid sequence with the ability to interact with the ribosome exit tunnel and arrest or at least slow its progression along the transcript. This idea will require future testing to validate.

Meanwhile, we also found numerous uORFs that are unlikely to encode translated peptides in multiple angiosperms. For instance, aside from having peptide encoding uORFs, both *bZIP53* and *SIP1* mRNAs have short non-peptide encoding uORFs in five angiosperms. Alignment of nucleotide sequences revealed that the regions flanking the start codons of these uORFs have the highest conservation level, suggesting that these non-peptide encoding uORFs probably function mainly via their start codons and flanking regions (Appendix S2B). Interestingly, consistently in seven angiosperms, *NIP5* mRNAs harbored a translating mini-uORF, consisting merely of a start and stop codon (Figure 9b). Although mini-uORFs were once thought to be untranslatable and had no biological function, the *NIP5* mini-uORF was first reported in Arabidopsis to function as a boron responsive element (Tanaka et al., 2016). Our results here indicated that *NIP5* uORF orthologs are also capable of being actively translated in another six angiosperm species in addition to Arabidopsis (Figure 9b). Moreover, we observed that the *NIP5* uORF is consistently a mini-uORF with well conserved flanking regions across all angiosperm orthologs (Figure 9c), implying that these flanking regions might function in facilitating the translation and regulatory functions of these sequences.

Although previous studies in plant uORFs mostly focused on the longer uORFs, our results indicate that it is also important to study the shorter uORFs, as they are conserved in multiple species, which implies important biological functions. In regard to functionality, although the functions of several abovementioned translating uORFs have been reported in Arabidopsis (Tanaka et al., 2016; Weltmeier et al., 2009; Wiese et al., 2004), it is unsure that if these uORFs serve the same function in other angiosperm species and among different conditions. Therefore, further characterization of the functions of uORFs in other species may be an important area of future inquiry.

In addition to finding the conserved nature of the abovementioned uORFs, we also identified several other never before predicted actively translating uORFs in the 5' UTRs of transcripts such as putative *eIF5*, *SAR1c*, and *LNK1* that are broadly conserved across the 10 studied angiosperms (Table 3). Overall, our results reveal that actively translating uORFs are highly prevalent with in the transcriptomes of various angiosperm species (Table 1), providing a resource for focusing future research endeavors on characterizing those that serve important regulatory functionality in numerous plant species. These future studies should not only focus on discovering the functions of novel uORFs we have identified but also on validating additional biological functionality of the uORFs that were reported previously.

Finally, we analyzed the sequence features of the regions flanking the uORF stop codon and identified conserved sequence features that might be related to the mechanisms of how these uORFs are able to stall ribosomes on angiosperm transcripts. Specifically, we found that the secondary structure immediately downstream of uORF stop

codons is more likely to be in a base-paired conformation compared to the regions that are further downstream or upstream of these sequences, and this phenomenon was consistent for uORFs that are actively translated in all the 10 angiosperm species (Figure 9d). According to previous studies in yeast, mRNA secondary structure can stall ribosomes during the elongation process (Doma & Parker, 2006; Schuller & Green, 2018). Therefore, it is possible that the secondary structures at the uORF stop codon we identified in these plant species contribute partially to stalling the ribosomes and damping mORF translation when these uORFs are active. However, additional experiments will be needed to validate if and how the secondary structures impact the uORF's ability to stall ribosomes in angiosperm transcriptomes.

Additionally, we analyzed the sequence context of the uORF stop codon flanking regions and found that certain nucleotides are overrepresented at specific locations relative to uORF flanking sequences. Interestingly, the nucleotides that we found to be overrepresented in these areas of uORF containing transcript 5' UTRs display monocot and dicot specific patterns (Figure 10). It has been found in human cells that certain nucleotides immediately after the stop codon can favor ribosome stop codon read through, suggesting that sequence context in the stop codon flanking regions can affect ribosome behavior (Loughran et al., 2014; Wangen & Green, 2020). Although it was difficult to tell how the overrepresented nucleotides identified in this study can contribute to uORF functionality in stalling ribosomes or some other mechanism, we speculate that these nucleotide sequences function by changing ribosome behavior when encountering a uORF stop codon. Future experiments will be needed to determine if and how these observed sequence elements that are monocot and dicot specific contribute to uORF-mediated regulation, either by ribosome stalling or some other mechanism.

Lastly, we analyzed the number of mRNAs harboring actively translating uORFs while being sorted into the co-translational decay pathway at the same time and discovered that actively translating uORFs and co-translational decay are almost entirely mutually exclusive in the majority of studied angiosperms (Figure 11 and Table 1), which implies that actively translating uORFs inhibit downstream co-translational decay. This result aligns with the previous studies because translating uORFs can arrest ribosomes and block downstream translation (Barbosa et al., 2013; Zhang et al., 2019), which is also likely to block co-translational decay (Figure 12b).

However, we did observe a higher percentage of mRNAs harboring translating uORFs and being co-translationally decayed at the same time in a specific Arabidopsis dataset and in common bean results (Figure 11 and Table 1). Given that it has been reported previously that translation in the main ORF can be reinitiated after uORF translation and sometimes ribosomes can even bypass a uORF via leaky scanning (Silva et al., 2019), the most likely explanation would be that many uORFs identified to be actively translating in these two datasets were not able to arrest ribosomes completely, allowing translation to be reinitiated in the main ORF, which then allows co-translational decay to occur.



However, more experiments are needed to test if this is the case. Together, our analyses revealed that actively translating uORFs are likely to inhibit downstream co-translational decay in most scenarios, and the inhibition could be related to the uORFs' ability to arrest ribosomes.

In total, we have confirmed that co-translational mRNA decay is a widespread and conserved process in angiosperms, and likely all plant species, as well as identified the sequence features associated with the co-translationally decayed mRNAs. Interestingly, we find that most of these sequence features are correlated with low translation efficiency, leading us to hypothesize that low translation efficiency is the major triggering mechanism of co-translational mRNA decay in angiosperms. We have also identified numerous actively translating uORFs across 10 angiosperm species using the same collection of degradome datasets and found conserved sequence features in the flanking regions of the uORFs' stop codons that can potentially contribute to stalling ribosomes at these translation regulatory elements. Meanwhile, we confirmed that actively translating uORFs inhibit downstream co-translational decay in most of the angiosperms. These findings provide further evidence that degradome datasets are rich resources for identifying and characterizing actively translating uORFs in eukaryotic species where ribosome profiling is not available or applicable. More importantly, our results provide important insights for future studies into the initiation mechanisms of co-translational mRNA decay and the ribosome stalling mechanisms of translating uORFs.

4 | MATERIALS AND METHODS

4.1 | Degradome data sources and preprocessing

The Arabidopsis genome (TAIR10) and gene annotation were downloaded from TAIR (<https://www.arabidopsis.org>). The genomes and gene annotations of tomato (*S. lycopersicum* iTAG2.4), soybean (*Glycine max* Wm82.a2.v1), Medicago (*M. truncatula* Mt4.0v1), common bean (*P. vulgaris* v2.1), rice (*Oryza sativa* v7_JGI), maize (*Z. mays* Ensembl-18), Setaria (*S. viridis* v2.1), sorghum (*S. bicolor* v3.1.1), and *Brachypodium* (*B. distachyon* v3.2) were downloaded from phytozome v12.1 (<https://phytozome.jgi.doe.gov/pz/portal.html>) (Goodstein et al., 2012). The coordinate of stop codons and mature mRNA sequences were extracted from the genomes based on the gene annotations. The raw data of plant degradome were retrieved from GEO dataset in NCBI (Table 1) (Goodstein et al., 2012). The coordinate of stop codons and mature mRNA sequences were extracted from the genomes based on the gene annotations. The raw data of plant degradome were retrieved from GEO dataset in NCBI (Table 1). Adaptors were trimmed using Cutadapt with default parameters (Martin, 2011). For PARE datasets, only reads that were 20–21 nt in length were used. The trimmed reads were mapped to the mature mRNA sequences (primary isoforms according to available annotations) using STAR tool with the parameters “--outFilterMultimapNmax 1

--outFilterMismatchNmax 1 --outFilterMismatchNoverLmax 0.10” (Dobin et al., 2013). The SAM files were further converted to BED files containing only the 5' most nucleotide of each read, denoting the 5'P intermediates. The 5'P read ends overlapping with the 100 nt flanking of stop codon were calculated using BEDTools (Quinlan & Hall, 2010). All the reads were normalized to reads per million (RPM). In order to minimize the noise in the sequence feature analysis, the mRNAs whose CDS length is not a multiple of three or contains unclear nucleotides (Ns) in its sequence, or the CDS do not end with UAA, UAG, or UGA were excluded from all the sequence feature analysis in order to avoid misannotated genes. Adaptors were trimmed using Cutadapt with default parameters (Martin, 2011). For PARE datasets, only reads that were 20–21 nt in length were used. The trimmed reads were mapped to the mature mRNA sequences (primary isoforms according to available annotations) using STAR tool with the parameters “--outFilterMultimapNmax 1 --outFilterMismatchNmax 1 --outFilterMismatchNoverLmax 0.10” (Dobin et al., 2013). The SAM files were further converted to BED files containing only the 5' most nucleotide of each read, denoting the 5'P intermediates. The 5'P read ends overlapping with the 100 nt flanking of stop codon were calculated using BEDTools (Quinlan & Hall, 2010). All the reads were normalized to reads per million (RPM). In order to minimize the noise in the sequence feature analysis, the mRNAs whose length is not a multiple of three or contains unclear nucleotides (Ns) in its sequence, or the CDS do not end with UAA, UAG, or UGA were excluded from all the sequence feature analysis in order to avoid misannotated genes.

4.2 | Identifying co-translationally decayed mRNAs

The stop codon coordinates of each main ORF were first obtained from the genome annotation for each angiosperm. The 5'P reads mapped to mature RNAs were overlapped with the flanking 100 nt of the stop codons for each main ORF using BEDTools (Quinlan & Hall, 2010). An mRNA was considered a CTRD target if its TSI was greater than three, and it had at least 20 raw reads in the flanking 100 nt of its main ORF stop codon. The heatmap of the scaled 5'P read end distributions in the flanking 100 nt of main ORF stop codon was plotted using a customized R script (Figure 1 and Figure S1).

4.3 | Codon usage analysis for co-translationally decayed mRNAs

$$RSCU = \frac{\text{Occurrence of a Specific Codon}}{\frac{1}{3} \times (\text{Total Occurrence of all Synonymous Codons})} \quad (\text{Sharp \& Li, 1986}).$$

n = total number of synonymous codons encoding an amino acid

$$\Delta RSCU = RSCU_{\text{CTRD}} - RSCU_{\text{non-CTRD}}$$

The statistical significance in RSCU difference for CTRD mRNAs and non-CTRD mRNAs was determined by adjusted student *t*-tests.

4.4 | Determining ribosome pausing at enriched codons and amino acids in co-translationally decayed mRNAs

The ribosome profiling datasets for Arabidopsis and tomato were retrieved from the GEO database. For each sample, reads from 28 to 30 nt were mapped to the corresponding genomes mentioned above. For each codon,

$$\text{Codon Stalling Index} = \frac{5' \text{P end reads}_{\text{A site}} + 5' \text{P end reads}_{\text{P site}}}{2 \times \text{median } 5' \text{P end reads of flanking } 100 \text{ nt}}$$

The A site and P site locations were calculated according to the corresponding papers (Hsu et al., 2016; Wu et al., 2019). For each amino acid,

$$\text{Amino Acid Stalling Index} = \frac{1}{n} \times \sum_{i=1}^n \text{Synonymous Codon Stalling Index}$$

n = total number of synonymous codons encoding an amino acid.

The statistical difference between synonymous codon stalling index was determined by students t -test.

4.5 | Prediction of putative uORFs and identifying those that are actively translating

All 5' UTR sequences from annotated mRNAs were extracted and scanned to identify potential ORFs that start with the canonical start codon ATG or near-cognate start codons ACG or CTG and end with a stop codon TAA, TAG, or TGA. To identify the predicted uORF candidates, the 5'P reads mapped to mature mRNAs were overlapped with the flanking 100 nt of potential stop codons of predicted uORFs using BEDTools (Quinlan & Hall, 2010). A uORF was considered as actively translating if its TSI was greater than three, and it had at least 20 raw reads in the flanking 100 nt of its stop codon.

4.6 | uORF sequence alignment

To assess the level of conservation of uORF sequences, the uORFs and their flanking sequence or the peptide encoded by the uORFs from each species were aligned with ClustalW linux version (<http://www.clustal.org/clustal2/>) (Larkin et al., 2007).

4.7 | Calculating the EFE of the regions flanking the uORF stop codons

Within the 100 nt regions flanking the stop codon of the identified actively translating uORFs in each species, the EFE was calculated using ViennaRNA Package using 35 nt sliding windows (Gruber et al., 2015).

4.8 | Sequence context analysis for the regions flanking uORF stop codons

The sequence contexts were generated by Seq2Logo Weighted Kullback–Leibler logo type using the 10 nt upstream of the uORF stop codon as input with a frequency matrix.

ACKNOWLEDGMENTS

The authors would like to thank the members of the B.D.G. lab both past and present for helpful discussions. This work has funded by NSF grants IOS-2023310 AND IOS-1849708 to B.D.G. The funders had no role in study design, literature collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest to report.

AUTHOR CONTRIBUTIONS

Brian D. Gregory, Rong Guo, and Xiang Yu conceived the study and designed the experiments. Rong Guo analyzed the data. Rong Guo and Brian D. Gregory wrote the paper. The authors have read and approved the manuscript for publication.

DATA AVAILABILITY STATEMENT

All data used in this study are cited in Table 1.

ORCID

Rong Guo  <https://orcid.org/0000-0002-7995-7889>

Xiang Yu  <https://orcid.org/0000-0002-5730-8802>

Brian D. Gregory  <https://orcid.org/0000-0001-7532-0138>

REFERENCES

- Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., & Axtell, M. J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Current Biology*, 18, 758–762. <https://doi.org/10.1016/j.cub.2008.04.042>
- Anderson, S. J., Kramer, M. C., Gosai, S. J., Yu, X., Vandivier, L. E., Nelson, A. D. L., Anderson, Z. D., Beilstein, M. A., Fray, R. G., Lyons, E., & Gregory, B. D. (2018). N(6)-Methyladenosine inhibits local Ribonucleolytic cleavage to stabilize mRNAs in Arabidopsis. *Cell Reports*, 25, 1146–1157.e1143. <https://doi.org/10.1016/j.celrep.2018.10.020>
- Baldrich, P., Campo, S., Wu, M. T., Liu, T. T., Hsing, Y. I., & San Segundo, B. (2015). MicroRNA-mediated regulation of gene expression in the response of rice plants to fungal elicitors. *RNA Biology*, 12, 847–863. <https://doi.org/10.1080/15476286.2015.1050577>
- Barbosa, C., Peixeiro, I., & Romão, L. (2013). Gene expression regulation by upstream open reading frames and human disease. *PLoS Genetics*, 9, e1003529. <https://doi.org/10.1371/journal.pgen.1003529>
- Behura, S. K., & Severson, D. W. (2013). Codon usage bias: Causative factors, quantification methods and genome-wide patterns: With emphasis on insect genomes. *Biological Reviews of the Cambridge Philosophical Society*, 88, 49–61. <https://doi.org/10.1111/j.1469-185X.2012.00242.x>
- Bhushan, S., Meyer, H., Starosta, A. L., Becker, T., Mielke, T., Berninghausen, O., Sattler, M., Wilson, D. N., & Beckmann, R. (2010). Structural basis for translational stalling by human cytomegalovirus



- and fungal arginine attenuator peptide. *Molecular Cell*, 40, 138–146. <https://doi.org/10.1016/j.molcel.2010.09.009>
- Branco-Price, C., Kawaguchi, R., Ferreira, R. B., & Bailey-Serres, J. (2005). Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation. *Annals of Botany*, 96, 647–660. <https://doi.org/10.1093/aob/mci217>
- Capella, M., Ribone, P. A., Arce, A. L., & Chan, R. L. (2015). Arabidopsis thaliana HomeoBox 1 (AtHB1), a Homeodomain-leucine zipper I (HD-zip I) transcription factor, is regulated by PHYTOCHROME-INTERACTING FACTOR 1 to promote hypocotyl elongation. *The New Phytologist*, 207, 669–682. <https://doi.org/10.1111/nph.13401>
- Chew, G. L., Pauli, A., & Schier, A. F. (2016). Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nature Communications*, 7, 11663. <https://doi.org/10.1038/ncomms11663>
- Crisp, P. A., Ganguly, D. R., Smith, A. B., Murray, K. D., Estavillo, G. M., Searle, I., Ford, E., Bogdanović, O., Lister, R., Borevitz, J. O., Eichten, S. R., & Pogson, B. J. (2017). Rapid recovery gene downregulation during excess-light stress and recovery in Arabidopsis. *Plant Cell*, 29, 1836–1863. <https://doi.org/10.1105/tpc.16.00828>
- De Coninck, B., Carron, D., Tavormina, P., Willem, L., Craik, D. J., Vos, C., Thevissen, K., Mathys, J., & Cammue, B. P. (2013). Mining the genome of Arabidopsis thaliana as a basis for the identification of novel bioactive peptides involved in oxidative stress tolerance. *Journal of Experimental Botany*, 64, 5297–5307. <https://doi.org/10.1093/jxb/ert295>
- Devers, E. A., Branscheid, A., May, P., & Krajinski, F. (2011). Stars and symbiosis: microRNA- and microRNA*-mediated transcript cleavage involved in arbuscular mycorrhizal symbiosis. *Plant Physiology*, 156, 1990–2010. <https://doi.org/10.1104/pp.111.172627>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Doma, M. K., & Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, 440, 561–564. <https://doi.org/10.1038/nature04530>
- Duncan, C. D. S., Rodríguez-López, M., Ruis, P., Bähler, J., & Mata, J. (2018). General amino acid control in fission yeast is regulated by a nonconserved transcription factor, with functions analogous to Gcn4/Atf4. *Proceedings of the National Academy of Sciences of the United States of America*, 115, E1829–e1838. <https://doi.org/10.1073/pnas.1713991115>
- Formey, D., Iñiguez, L. P., Peláez, P., Li, Y. F., Sunkar, R., Sánchez, F., Reyes, J. L., & Hernández, G. (2015). Genome-wide identification of the *Phaseolus vulgaris* sRNAome using small RNA and degradome sequencing. *BMC Genomics*, 16, 423. <https://doi.org/10.1186/s12864-015-1639-5>
- Franke, K. R., Schmidt, S. A., Park, S., Jeong, D. H., Accerbi, M., & Green, P. J. (2018). Analysis of *Brachypodium* miRNA targets: Evidence for diverse control during stress and conservation in bioenergy crops. *BMC Genomics*, 19, 547. <https://doi.org/10.1186/s12864-018-4911-7>
- Gerashchenko, M. V., Lobanov, A. V., & Gladyshev, V. N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 17394–17399. <https://doi.org/10.1073/pnas.1120799109>
- German, M. A., Luo, S., Schroth, G., Meyers, B. C., & Green, P. J. (2009). Construction of parallel analysis of RNA ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nature Protocols*, 4, 356–362. <https://doi.org/10.1038/nprot.2009.8>
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, 40, D1178–D1186. <https://doi.org/10.1093/nar/gkr944>
- Gregory, B. D., O'Malley, R. C., Lister, R., Urlich, M. A., Tonti-Filippini, J., Chen, H., Millar, A. H., & Ecker, J. R. (2008). A link between RNA metabolism and silencing affecting Arabidopsis development. *Developmental Cell*, 14, 854–866. <https://doi.org/10.1016/j.devcel.2008.04.005>
- Gruber, A. R., Bernhart, S. H., & Lorenz, R. (2015). The ViennaRNA web services. *Methods in Molecular Biology*, 1269, 307–326. https://doi.org/10.1007/978-1-4939-2291-8_19
- Hayden, C. A., & Jorgensen, R. A. (2007). Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. *BMC Biology*, 5, 32. <https://doi.org/10.1186/1741-7007-5-32>
- Hou, C. Y., Lee, W. C., Chou, H. C., Chen, A. P., Chou, S. J., & Chen, H. M. (2016). Global analysis of truncated RNA ends reveals new insights into ribosome stalling in plants. *Plant Cell*, 28, 2398–2416. <https://doi.org/10.1105/tpc.16.00295>
- Hsu, P. Y., & Benfey, P. N. (2018). Small but mighty: Functional peptides encoded by small ORFs in plants. *Proteomics*, 18, e1700038. <https://doi.org/10.1002/pmic.201700038>
- Hsu, P. Y., Calviello, L., Wu, H. L., Li, F. W., Rothfels, C. J., Ohler, U., & Benfey, P. N. (2016). Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 113, E7126–e7135. <https://doi.org/10.1073/pnas.1614788113>
- Hu, W., Petzold, C., Collier, J., & Baker, K. E. (2010). Nonsense-mediated mRNA decapping occurs on polyribosomes in *Saccharomyces cerevisiae*. *Nature Structural & Molecular Biology*, 17, 244–247. <https://doi.org/10.1038/nsmb.1734>
- Hu, W., Sweet, T. J., Chamnongpol, S., Baker, K. E., & Collier, J. (2009). Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature*, 461, 225–229. <https://doi.org/10.1038/nature08265>
- Ibrahim, F., Maragkakis, M., Alexiou, P., & Mourelatos, Z. (2018). Ribothrypsis, a novel process of canonical mRNA decay, mediates ribosome-phased mRNA endonucleolysis. *Nature Structural & Molecular Biology*, 25, 302–310. <https://doi.org/10.1038/s41594-018-0042-8>
- Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., & Weissman, J. S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*, 7, 1534–1550. <https://doi.org/10.1038/nprot.2012.086>
- Jeong, D. H., Schmidt, S. A., Rymarquis, L. A., Park, S., Ganssmann, M., German, M. A., Accerbi, M., Zhai, J., Fahlgren, N., Fox, S. E., Garvin, D. F., Mockler, T. C., Carrington, J. C., Meyers, B. C., & Green, P. J. (2013). Parallel analysis of RNA ends enhances global investigation of microRNAs and target RNAs of *Brachypodium distachyon*. *Genome Biology*, 14, R145. <https://doi.org/10.1186/gb-2013-14-12-r145>
- Jiao, Y., & Meyerowitz, E. M. (2010). Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Molecular Systems Biology*, 6, 419. <https://doi.org/10.1038/msb.2010.76>
- Kawaguchi, R., & Bailey-Serres, J. (2005). mRNA sequence features that contribute to translational regulation in Arabidopsis. *Nucleic Acids Research*, 33, 955–965. <https://doi.org/10.1093/nar/gki240>
- Laing, W. A., Martínez-Sánchez, M., Wright, M. A., Bulley, S. M., Brewster, D., Dare, A. P., Rassam, M., Wang, D., Storey, R., Macknight, R. C., & Hellens, R. P. (2015). An upstream open reading frame is essential for feedback regulation of ascorbate biosynthesis in Arabidopsis. *Plant Cell*, 27, 772–786. <https://doi.org/10.1105/tpc.114.133777>
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and

- Clustal X version 2.0. *Bioinformatics*, 23, 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Lei, L., Shi, J., Chen, J., Zhang, M., Sun, S., Xie, S., Li, X., Zeng, B., Peng, L., Hauck, A., Zhao, H., Song, W., Fan, Z., & Lai, J. (2015). Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *The Plant Journal*, 84, 1206–1218. <https://doi.org/10.1111/tpj.13073>
- Li, Y. F., Zheng, Y., Addo-Quaye, C., Zhang, L., Saini, A., Jagadeeswaran, G., Axtell, M. J., Zhang, W., & Sunkar, R. (2010). Transcriptome-wide identification of microRNA targets in rice. *The Plant Journal*, 62, 742–759. <https://doi.org/10.1111/j.1365-3113X.2010.04187.x>
- Liu, H., Qin, C., Chen, Z., Zuo, T., Yang, X., Zhou, H., Xu, M., Cao, S., Shen, Y., Lin, H., He, X., Zhang, Y., Li, L., Ding, H., Lübberstedt, T., Zhang, Z., & Pan, G. (2014). Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. *BMC Genomics*, 15, 25. <https://doi.org/10.1186/1471-2164-15-25>
- Liu, M. J., Wu, S. H., Wu, J. F., Lin, W. D., Wu, Y. C., Tsai, T. Y., Tsai, H. L., & Wu, S. H. (2013). Translational landscape of photomorphogenic *Arabidopsis*. *Plant Cell*, 25, 3699–3710. <https://doi.org/10.1105/tpc.113.114769>
- Loughran, G., Chou, M. Y., Ivanov, I. P., Jungreis, I., Kellis, M., Kiran, A. M., Baranov, P. V., & Atkins, J. F. (2014). Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Research*, 42, 8928–8938. <https://doi.org/10.1093/nar/gku608>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Nagarajan, V. K., Jones, C. I., Newbury, S. F., & Green, P. J. (2013). XRN 5'→3' exoribonucleases: Structure, mechanisms and functions. *Biochimica et Biophysica Acta*, 1829, 590–603. <https://doi.org/10.1016/j.bbagr.2013.03.005>
- Parker, R. (2012). RNA degradation in *Saccharomyces cerevisiae*. *Genetics*, 191, 671–702. <https://doi.org/10.1534/genetics.111.137265>
- Pelechano, V., Wei, W., & Steinmetz, L. M. (2015). Widespread co-translational RNA decay reveals ribosome dynamics. *Cell*, 161, 1400–1412. <https://doi.org/10.1016/j.cell.2015.05.008>
- Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E., Graveley, B. R., & Collier, J. (2015). Codon optimality is a major determinant of mRNA stability. *Cell*, 160, 1111–1124. <https://doi.org/10.1016/j.cell.2015.02.029>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Ribone, P. A., Capella, M., Arce, A. L., & Chan, R. L. (2017). A uORF represses the transcription factor AtHB1 in aerial tissues to avoid a deleterious phenotype. *Plant Physiology*, 175, 1238–1253. <https://doi.org/10.1104/pp.17.01060>
- Rymarquis, L. A., Souret, F. F., & Green, P. J. (2011). Evidence that XRN4, an *Arabidopsis* homolog of exoribonuclease XRN1, preferentially impacts transcripts with certain sequences or in particular functional categories. *RNA*, 17, 501–511. <https://doi.org/10.1261/rna.2467911>
- Sabi, R., & Tuller, T. (2015). A comparative genomics study on the effect of individual amino acids on ribosome stalling. *BMC Genomics*, 16(Suppl 10), S5. <https://doi.org/10.1186/1471-2164-16-S10-S5>
- Sachs, M. S., & Geballe, A. P. (2006). Downstream control of upstream open reading frames. *Genes & Development*, 20, 915–921. <https://doi.org/10.1101/gad.1427006>
- Schoenberg, D. R., & Maquat, L. E. (2012). Regulation of cytoplasmic mRNA decay. *Nature Reviews. Genetics*, 13, 246–259. <https://doi.org/10.1038/nrg3160>
- Schuller, A. P., & Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. *Nature Reviews. Molecular Cell Biology*, 19, 526–541. <https://doi.org/10.1038/s41580-018-0011-4>
- Shamimuzzaman, M., & Vodkin, L. (2012). Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. *BMC Genomics*, 13, 310. <https://doi.org/10.1186/1471-2164-13-310>
- Shamimuzzaman, M., & Vodkin, L. (2018). Ribosome profiling reveals changes in translational status of soybean transcripts during immature cotyledon development. *PLoS ONE*, 13, e0194596. <https://doi.org/10.1371/journal.pone.0194596>
- Sharp, P. M., & Li, W. H. (1986). Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Research*, 14, 7737–7749. <https://doi.org/10.1093/nar/14.19.7737>
- Silva, J., Fernandes, R., & Romão, L. (2019). Translational regulation by upstream open reading frames and human diseases. *Advances in Experimental Medicine and Biology*, 1157, 99–116. https://doi.org/10.1007/978-3-030-19966-1_5
- Song, Q. X., Liu, Y. F., Hu, X. Y., Zhang, W. K., Ma, B., Chen, S. Y., & Zhang, J. S. (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biology*, 11, 5. <https://doi.org/10.1186/1471-2229-11-5>
- Takahashi, H., Takahashi, A., Naito, S., & Onouchi, H. (2012). BAIUCAS: A novel BLAST-based algorithm for the identification of upstream open reading frames with conserved amino acid sequences and its application to the *Arabidopsis thaliana* genome. *Bioinformatics*, 28, 2231–2241. <https://doi.org/10.1093/bioinformatics/bts303>
- Tanaka, M., Sotta, N., Yamazumi, Y., Yamashita, Y., Miwa, K., Murota, K., Chiba, Y., Hirai, M. Y., Akiyama, T., Onouchi, H., Naito, S., & Fujiwara, T. (2016). The minimum open reading frame, AUG-stop, induces boron-dependent ribosome stalling and mRNA degradation. *Plant Cell*, 28, 2830–2849. <https://doi.org/10.1105/tpc.16.00481>
- Tanguay, R. L., & Gallie, D. R. (1996). Translational efficiency is regulated by the length of the 3' untranslated region. *Molecular and Cellular Biology*, 16, 146–156. <https://doi.org/10.1128/MCB.16.1.146>
- Tran, M. K., Schultz, C. J., & Baumann, U. (2008). Conserved upstream open reading frames in higher plants. *BMC Genomics*, 9, 361. <https://doi.org/10.1186/1471-2164-9-361>
- Uchiyama-Kadokura, N., Murakami, K., Takemoto, M., Koyanagi, N., Murota, K., Naito, S., & Onouchi, H. (2014). Polyamine-responsive ribosomal arrest at the stop codon of an upstream open reading frame of the AdoMetDC1 gene triggers nonsense-mediated mRNA decay in *Arabidopsis thaliana*. *Plant & Cell Physiology*, 55, 1556–1567. <https://doi.org/10.1093/pcp/pcu086>
- Vaughn, J. N., Ellingson, S. R., Mignone, F., & Arim, A. (2012). Known and novel post-transcriptional regulatory sequences are conserved across plant families. *RNA*, 18, 368–384. <https://doi.org/10.1261/rna.031179.111>
- Wallace, E. W. J., Maufrais, C., Sales-Lee, J., Tuck, L. R., de Oliveira, L., Feuerbach, F., Moyrand, F., Natarajan, P., Madhani, H. D., & Janbon, G. (2020). Quantitative global studies reveal differential translational control by start codon context across the fungal kingdom. *Nucleic Acids Research*, 48, 2312–2331. <https://doi.org/10.1093/nar/gkaa060>
- Wangen, J. R., & Green, R. (2020). Stop codon context influences genome-wide stimulation of termination codon readthrough by aminoglycosides. *eLife*, 9, e52611. <https://doi.org/10.7554/eLife.52611>
- Weltmeier, F., Rahmani, F., Ehlert, A., Dietrich, K., Schütze, K., Wang, X., Chaban, C., Hanson, J., Teige, M., Harter, K., Vicente-Carbajosa, J., Smeekens, S., & Dröge-Laser, W. (2009). Expression patterns within the *Arabidopsis* C/S1 bZIP transcription factor network: Availability of heterodimerization partners controls gene



- expression during stress response and development. *Plant Molecular Biology*, 69, 107–119. <https://doi.org/10.1007/s11103-008-9410-9>
- Whittle, C. A., Sun, Y., & Johannesson, H. (2011). Evolution of synonymous codon usage in *Neurospora tetrasperma* and *Neurospora discreta*. *Genome Biology and Evolution*, 3, 332–343. <https://doi.org/10.1093/gbe/evr018>
- Wiese, A., Elzinga, N., Wobbes, B., & Smekens, S. (2004). A conserved upstream open reading frame mediates sucrose-induced repression of translation. *Plant Cell*, 16, 1717–1729. <https://doi.org/10.1105/tpc.019349>
- Wu, H. L., Song, G., Walley, J. W., & Hsu, P. Y. (2019). The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. *Plant Physiology*, 181, 367–380. <https://doi.org/10.1104/pp.19.00541>
- Yan, X., Hoek, T. A., Vale, R. D., & Tanenbaum, M. E. (2016). Dynamics of translation of single mRNA molecules in vivo. *Cell*, 165, 976–989. <https://doi.org/10.1016/j.cell.2016.04.034>
- Yu, X., Willmann, M. R., Anderson, S. J., & Gregory, B. D. (2016). Genome-wide mapping of uncapped and cleaved transcripts reveals a role for the nuclear mRNA cap-binding complex in cotranslational RNA decay in *Arabidopsis*. *Plant Cell*, 28, 2385–2397. <https://doi.org/10.1105/tpc.16.00456>
- Zhang, H., Wang, Y., & Lu, J. (2019). Function and evolution of upstream ORFs in eukaryotes. *Trends in Biochemical Sciences*, 44, 782–794. <https://doi.org/10.1016/j.tibs.2019.03.002>
- Zhao, D., Hamilton, J. P., Hardigan, M., Yin, D., He, T., Vaillancourt, B., Reynoso, M., Pauluzzi, G., Funkhouser, S., Cui, Y., Bailey-Serres, J., Jiang, J., Buell, C. R., & Jiang, N. (2017). Analysis of ribosome-associated mRNAs in rice reveals the importance of transcript size and GC content in translation. *G3 (Bethesda)*, 7, 203–219. <https://doi.org/10.1534/g3.116.036020>
- Zhou, M., Gu, L., Li, P., Song, X., Wei, L., Chen, Z., & Cao, X. (2010). Degradome sequencing reveals endogenous small RNA targets in rice (*Oryza sativa* L. ssp. indica). *Frontiers in Biology*, 5, 67–90. <https://doi.org/10.1007/s11515-010-0007-8>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Guo, R., Yu, X., & Gregory, B. D. (2023). The identification of conserved sequence features of co-translationally decayed mRNAs and upstream open reading frames in angiosperm transcriptomes. *Plant Direct*, 7(1), e479. <https://doi.org/10.1002/pld3.479>