

Upstream open reading frames buffer translational variability during *Drosophila* evolution and development

Reviewed Preprint

v1 • January 6, 2025

Not revised

Yuanqiang Sun, Yuange Duan, Peixiang Gao, Chenlu Liu, Kaichun Jin, Shengqian Dou, Wenxiong Tang, Hong Zhang, Jian Lu 

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China • Eye Institute of Shandong First Medical University, State Key Laboratory Cultivation Base, Shandong Provincial Key Laboratory of Ophthalmology, Qingdao, China • College of Ecology, Lanzhou University, Lanzhou, China

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

eLife Assessment

This study reveals the **important** role of upstream open reading frames (uORFs) in limiting the translational variability of downstream coding sequences. Through a combination of computational simulations, comparative analyses of translation efficiency across different developmental stages in two closely related *Drosophila* species, and manipulative, experimental validation of translation buffering by an uORF for a gene, the authors provide **convincing** evidence supporting their conclusions. This work will be of broad interest to molecular biologists and geneticists.

<https://doi.org/10.7554/eLife.104074.1.sa2>

Abstract

Protein abundance tends to be more evolutionarily conserved than mRNA levels both within and between species, yet the mechanisms underlying this phenomenon remain largely unknown. Upstream open reading frames (uORFs) are widespread *cis*-regulatory elements in eukaryotic genomes that regulate translation, but it remains unclear whether and how uORFs contribute to stabilizing protein levels. In this study, we performed ribosome translation simulations on mRNA to quantitatively assess the extent to which uORF translation influences the translational variability of downstream coding sequences (CDS) across varying contexts. Our simulations revealed that uORF translation dampens CDS translational variability, with buffering capacity increasing in proportion to uORF efficiency, length, and number. We then compared the translomes at different developmental stages of two *Drosophila* species, demonstrating that uORFs buffer mRNA translation fluctuations during both evolution and development. Experimentally, deleting a uORF in the *bcd* gene—a prominent example of translational buffering—resulted in extensive changes in gene expression and phenotypes in *Drosophila melanogaster*. Additionally, we observed uORF-mediated buffering between primates and within human populations. Together, our results

reveal a novel regulatory mechanism by which uORFs stabilize gene translation during development and across evolutionary time.

Introduction

Organisms have evolved various strategies for the spatiotemporal regulation of gene expression ^{1–3}. This is important because aberrant gene expression can result in phenotypic defects or diseases, while the variation and evolution of gene expression patterns frequently promote phenotypic diversification and adaptation ⁴. Although variations in mRNA abundance are widely observed within or between species, protein abundance tends to show stronger evolutionary constraint ^{5,6}, as observed in yeasts ^{7–9}, primates ^{10,11}, and other organisms ^{12–14}. Nevertheless, the molecular mechanisms by which the conservation of protein abundance across species is achieved are largely unknown ^{5,6}.

Eukaryotic mRNA translation is a crucial step in gene expression and is highly regulated by multilayered mechanisms ^{15–17}. Upstream open reading frames (uORFs), short open reading frames in the 5-terminal untranslated regions (5'UTRs) of eukaryotic mRNAs, play crucial roles in regulating mRNA translation. Approximately 50% of eukaryotic genes contain uORFs ¹⁸, and their evolution has been tightly shaped by natural selection ^{19–23}. The functions of uORFs have been explored in various contexts, including development ^{23–30}, disease ^{31–35} and stress responses ^{36–43}. The prevailing consensus is that uORFs typically repress downstream coding sequence (CDS) translation by sequestering ribosomes, a process influenced by factors such as uORF length, position, and sequence context ^{34,43–48}. However, under stress conditions, certain uORFs can facilitate CDS translation by promoting ribosome reinitiation, illustrating their context-dependent functions ^{42,43,48–52}.

Gene expression noise, which arises from the inherent stochasticity of biological processes such as transcription and translation, is generally detrimental to organismal fitness ⁵³ and is primarily determined at the translational level ⁵⁴. Recent studies suggest uORFs might play essential roles in buffering translational noise and stabilizing protein expression. For example, Wu et al. (2022) demonstrated that uORFs reduce protein production rates to stabilize TOC1 protein levels, ensuring precise circadian clock function in plants ⁵⁵. Similarly, Bottorff et al. (2022) used a human cell reporter system to show that a ribosome stall in cytomegaloviral *UL4* uORFs buffers against CDS translation reductions ⁵⁶. Under stress, translation initiation is typically downregulated, yet most human mRNAs resistant to this inhibition contain translated uORFs, with a single uORF often being sufficient for resistance ⁴². The computational model of Initiation Complexes Interference with Elongating Ribosomes (ICIER) suggests that derepression of downstream translation is a general mechanism of uORF-mediated stress resistance ⁴³. Despite these findings, the current understanding of uORFs in stabilizing translation is limited to single-gene cases or stressed conditions. It remains unclear whether and how uORFs affect gene translation variability on a genome-wide scale during evolution and development, and whether the identified mechanisms are universal or vary significantly among different taxa. To address these questions, a combination of modeling, genome-wide analyses, and comparative studies across species is required.

In this study, we first adapted the ICIER framework ⁴³ to simulate the translating ribosome on a mRNA to quantitatively measure the extent to which uORF translation reduces the translational variability of the downstream CDS under different translation contexts. We then compared the translomes of two closely related *Drosophila* species, *D. melanogaster* and *D. simulans*, and further supported the notion that uORFs could buffer the fluctuations of CDS translation during the development and evolution of *Drosophila*. The patterns also reappeared among primates and human populations. We next knocked out the *bcd* uORF, a case showing significant buffering effect

in our data, and observed wide changes in embryonic transcriptome and phenotypic defects in *D. melanogaster*. Together, our results demonstrate a novel role for uORFs by maintaining translation stabilization during *Drosophila* evolution and development.

Results

An extended ICIER model for quantifying uORF buffering in CDS translation

To quantitatively assess how uORF translation modulates the variability of downstream CDS translation, we adapted the ICIER model⁴³, originally grounded in the totally asymmetric simple exclusion process (TASEP). TASEP has been extensively utilized to model the stochastic nature of ribosome movement along mRNA, capturing the effects of ribosome traffic jams, where ribosomes may slow down or stall when a site ahead is occupied^{57–60}. The ICIER model extends this by simulating the interplay between scanning (40S) and elongating (80S) ribosomes, particularly focusing on how uORFs impact the overall translation efficiency of the main CDS⁴³.

We extended the original ICIER model⁴³ with several major modifications (**Fig. 1A**). First, while the original ICIER model only considered the scenario where the elongating ribosome (80S) causes downstream scanning ribosomes (40S) to dissociate from the mRNA when they move along the mRNA and collide, recent findings have shown that upstream dissociation can also play a critical role in uORF-mediated regulation⁵⁶. To incorporate this, we accounted for more complex ribosome interactions, including three possible scenarios where the 80S collides with 40S: i) 80S only causes the downstream 40S to dissociate from the mRNA with a probability of K_{down} (“downstream dissociation”, K_{down} ranging from 0 to 1), following the original ICIER model; ii) the 80S only causes the upstream 40S to dissociate from the mRNA with a probability of K_{up} (“upstream dissociation”, $K_{up} > 0$ and $K_{down} = 0$); and iii) a combination of the downstream and upstream dissociation models (“double dissociation”, $K_{up} > 0$ and $K_{down} > 0$).

Second, the original ICIER model only considered the ribosome collision and dissociation in the uORF and counted the 40S scanning ribosome escaping from the uORF as a proxy of the CDS translation rate. In our extended model, we also considered the ribosome collision and dissociation events in the CDS that is downstream of the uORF. We recorded the number of 80S ribosomes that completed translation at the stop codon of a CDS (N_{EC}) or uORF (N_{Eu}) during a given time interval, using these counts as proxies to quantify the translation efficiency (TE) of CDS or uORF. These indices allowed us to directly and quantitatively measure the impact of uORF-mediated translational buffering.

Third, while the original ICIER model only considered the effect of a single uORF, we also considered the buffering effects of two uORFs, allowing us to test the possible combinatorial effects of uORF-mediated translation regulation in this more complex yet more common scenario, as previous studies have shown uORFs tend to be clustered in genes¹⁸.

These extensions allow for a more comprehensive exploration of uORF-mediated translational buffering, offering deeper insights into how these regulatory elements might stabilize protein synthesis across varying translation contexts.

uORF-mediated buffering of CDS translation across different parameter settings

To systematically investigate the extent to which uORF translation modulates the variability of downstream CDS translation, we conducted simulations across various parameter settings using three dissociation models: upstream, downstream, and double dissociation, each reflecting

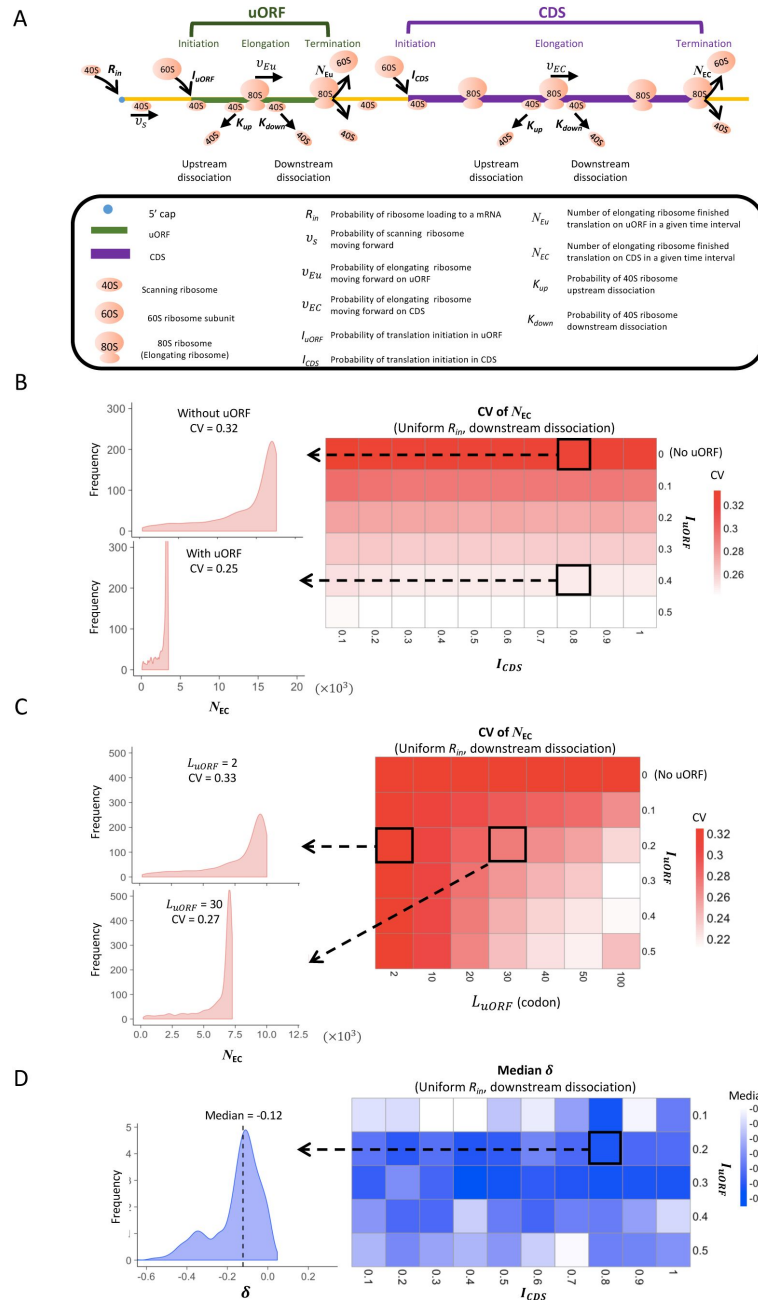


Figure 1.

Modeling simulation of uORF-mediated translation buffering.

(A) Model schema of the modified ICIER model (on the top); the parameters are listed in the box below the schema. (B) Heatmap showing the CVs of CDS TE (N_{EC}) under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. The left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under $I_{CDS} = 0.8$ & $I_{uORF} = 0$ (top panel, without uORF) and $I_{CDS} = 0.8$ & $I_{uORF} = 0.4$ (bottom panel, with uORF). (C) Heatmap showing CVs of CDS TE (N_{EC}) under different I_{uORF} (x-axis) and L_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. The left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under $L_{uORF} = 2$ & $I_{uORF} = 0.2$ (top panel) and $L_{uORF} = 30$ & $I_{uORF} = 0.2$ (bottom panel). (D) Heatmap showing median δ [$\log_2(\Delta N_{EC}/\Delta N_{EU})$] under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. The left panel elicited by the dotted line from a specific square of right heatmap was an example showing the distribution of δ under $I_{CDS} = 0.8$ & $I_{uORF} = 0.2$. The vertical dashed line indicated the median value of δ .

different possible interactions between scanning and elongating ribosomes on mRNA. We considered a range of parameters crucial to the translation process (Table S1), including the length of the 5' leader before the uORF (fixed at 150 nucleotides), the length of the uORF itself (ranging from 2 to 100 codons), the distance between the uORF stop codon and the CDS start codon (150 nucleotides), the length of the CDS (500 codons), and the length of the 3' UTR (150 nucleotides). Additionally, we modeled the probabilities associated with ribosome movement and initiation, such as the probability of a 40S ribosome moving to the next nucleotide ($v_s = 0.3$), the probability of an 80S ribosome moving to the next position within the uORF ($v_{Eu} = 0.3$) and within the CDS ($v_{EC} = 0.5$), and the probability of loading a new 40S ribosome at the 5' end of the mRNA (R_{in}). We also explored different probabilities of translation initiation at both the uORF start codon (I_{uORF}) and the CDS start codon (I_{CDS}). Key parameters were adapted from the original ICIER model⁴³, ensuring a robust basis for comparison while allowing exploration of additional variables that influence uORF-mediated translational buffering.

In our simulations, R_{in} values were varied to simulate fluctuations in translational resources, such as ribosome availability, that could arise from genetic differences or environmental changes during evolution or development. By generating 1,000 R_{in} values following either uniform or exponential distributions (ranging from 0 to 0.1, Fig. S1A), we aimed to capture the natural variability in ribosome loading rates that might occur across different species, individuals, or developmental stages. These values were then fed into the simulation models to evaluate their impact on both the level and variability of translation efficiency for CDSs, with the number of ribosomes completing translation on CDSs (N_{EC}) in a given time interval used as proxies for translational efficiency (Fig. S1B).

Across all model settings, uORF translation ($I_{uORF} > 0$) consistently reduced CDS translation efficiency (N_{EC}) by about 30% to 80% as I_{uORF} increased from 0.1 to 0.5, compared to scenarios where the uORF was absent or untranslated ($I_{uORF} = 0$) (Fig. S2 and S3). This confirms the inhibitory effect of uORF on downstream CDS translation. Notably, the coefficient of variation (CV), a measurement of variability for translation efficiency (N_{EC}), was lower when the uORF was translated (Fig. 1B). These CV values further decreased by approximately 10% to 25% as I_{uORF} increased from 0.1 to 0.5 (Fig. 1B), and this buffering effect persisted across different parameter settings (Fig. S4 and S5). Moreover, the CV of translation efficiency (N_{EC}) further decreased by about 6% to 30% as uORF length (I_{uORF}) increased from 2 to 100 codons (Fig. 1C and S6-7). These simulation results indicate that uORF translation can reduce variability in downstream CDS translation, with the buffering capacity positively correlated with both uORF translation initiation efficiency and length under the applied simulation conditions.

To more quantitatively investigate the relationship between changes in uORF (N_{EU}) and CDS (N_{EC}) translation, for each parameter setting, we used the median value (R_{in2}) of the 1000 R_{in} inputs as the baseline, calculating the corresponding N_{EC2} and N_{EU2} values. We then calculated the difference in changes in N_{EC} (ΔN_{EC}) and N_{EU} (ΔN_{EU}) relative to N_{EC2} and N_{EU2} for each R_{in} value. Across different parameter settings, ΔN_{EU} was consistently and significantly positively correlated with ΔN_{EC} ($P < 0.001$, Spearman's correlation) (Fig. S8-9), indicating that fluctuations of the ribosome loading rate influence the translation of both uORFs and CDSs in the same direction. Nevertheless, our simulations showed that variations in R_{in} led to a larger change in N_{EU} than in N_{EC} , as the median value of δ [defined by $\log_2(\Delta N_{EC}/\Delta N_{EU})$] was consistently less than 0 across various I_{uORF} and I_{CDS} combinations (Fig. 1D and Fig. S10-11). This finding suggests that uORFs buffer against upstream fluctuations, exhibiting greater translational fluctuations than downstream CDSs.

Simulations of mRNAs with two uORFs revealed patterns consistent with a buffering role for uORFs (Fig. S12). To compare the buffering effects of a single uORF versus two uORFs, we calculated the ratio of the CV of N_{EC} with two uORFs to that with a single uORF. A ratio less than 1 suggests that two uORFs provide greater buffering than a single uORF. For comparability, we

examined the CV of N_{EC} where the I_{uORF} in the single-uORF model equals $I_{uORF=}$ (I_{uORF} of the first uORF) in the two-uORF model, both ranging from 0 to 0.5. The CV ratio consistently remained below 1 across a range of $I_{uORF>}$ (I_{uORF} of the second uORF) (Fig. S13), indicating that two uORFs offer stronger buffering than a single uORF.

Collectively, these simulations collectively suggest: (1) uORF-mediated translational control buffers against CDS translation variability, (2) uORFs exhibit greater translational fluctuations than downstream CDSs, and (3) the buffering capacity of uORFs positively correlates with their translation initiation efficiency and length. Subsequently, we sought to validate these simulation results in a biological context by confirming the uORF-mediated buffering effect during organismal evolution and development, as both processes face environmental and/or genetic changes that frequently disturb mRNA translation.

Generating matched transcriptome data from two *Drosophila* species for comparative analysis

To validate the uORF-mediated translational buffering during *Drosophila* evolution, we performed a comparative analysis of the transcriptomes of two closely related *Drosophila* species, *D. melanogaster* and *D. simulans*, which diverged approximately 5.4 million years ago⁶¹. We generated high-throughput sequencing data for *D. simulans*, including transcriptome (mRNA-Seq) and transcriptome (Ribo-Seq) profiles from various developmental stages and tissues, such as embryos at 0-2 h, 2-6 h, 6-12 h, and 12-24 h, third-instar larvae, P7-8 pupae, female and male bodies, and female and male heads. In total, we obtained approximately 786 million high-quality reads for *D. simulans* (Table S2). These datasets were designed to be directly comparable to the previously published *D. melanogaster* data²³, with identical embryonic stages and tissue types. This comprehensive comparative transcriptome analysis, utilizing matched developmental stages and tissues between the two species, allowed us to test the uORF-mediated translational buffering effects during evolution.

Translational conservation and dominance of uORFs between *Drosophila* species

Given that the translation of uORFs is a crucial determinant for their functional impact, we first characterized the translational profiles of uORFs in *D. melanogaster* and *D. simulans* to explore the evolutionary roles of uORFs in gene regulation. We identified 18,412 canonical uORFs shared between the two species (referred to as conserved uORFs hereafter), 2,789 uORFs specific to *D. melanogaster* and 2,440 uORFs specific to *D. simulans*. The translational efficiencies (TEs) of the conserved uORFs were highly correlated between the two species across all developmental stages and tissues examined, with Spearman correlation coefficients ranging from 0.478 to 0.573 (Fig. 2A). Notably, conserved uORFs exhibited significantly higher translational efficiencies (TEs) compared to species-specific uORFs in both species. The median TE of conserved uORFs was 1.62 times that of non-conserved uORFs in *D. simulans*, while the corresponding ratio in *D. melanogaster* was 1.52 (Fig. 2B).

In *D. melanogaster*, 7,259 (52.2%) genes had no uORFs, 2,687 (19.3%) had a single uORF, and 3,961 (28.5%) contained multiple uORFs. Among genes with multiple uORFs, one uORF generally emerged as dominant, displaying a higher TE than the others within the same gene (Fig. 2C). The median TE of the dominant uORF was 4.84 times that of the second-highest uORF within the same gene in *D. melanogaster*, and the corresponding ratio was 5.21 times in *D. simulans* (Fig. 2C). To assess the consistency of this dominance across different tissues and developmental stages, we identified 3,072 multiple-uORF genes in *D. melanogaster* with at least one translated uORF (TE > 0.1 in at least five stages/tissues). Of these, 569 genes consistently used the same dominant uORF across the measured samples, significantly higher than the number expected under randomness (5 genes, 95% confidence interval: 1-10) based on shuffling the TEs of uORFs

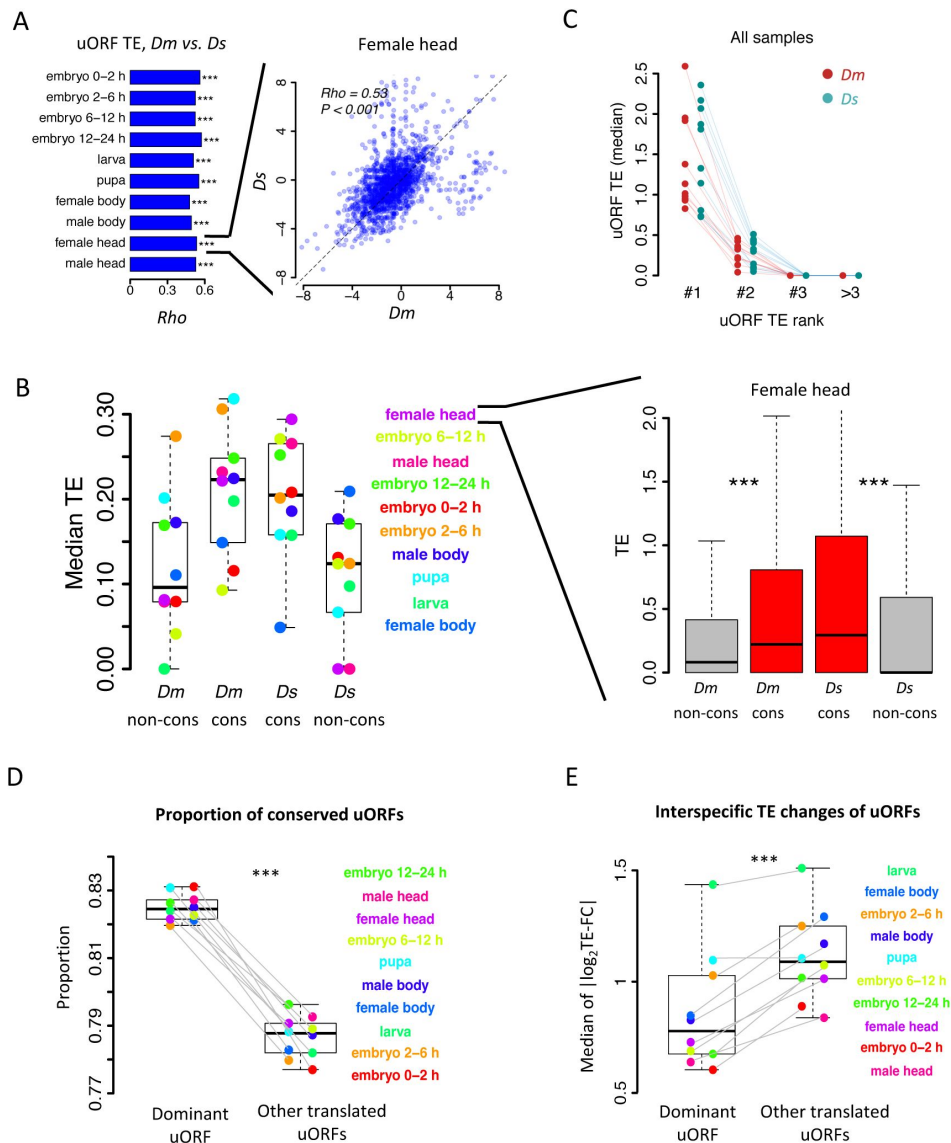


Figure 2.

Conservation and translation of uORFs between *D. melanogaster* and *D. simulans*.

(A) Spearman's correlation coefficients (Rho , represented by the bars) of conserved TEs between Dm (*D. melanogaster*) and Ds (*D. simulans*). ***, $P < 0.001$. Data for the female head sample is shown as an example in the right panel. The x- and y-axes represent the uORF TEs in Dm and Ds. (B) The median of TE of conserved and species-specific uORFs in each sample. Each dot represents the median TE of a sample for a specific uORF class. Data from the female head sample is shown as an example in the right panel. P values were obtained from Wilcoxon rank sum tests. ***, $P < 0.001$. (C) uORFs were ranked by decreasing TEs within each gene. The uORF with the highest TE within each gene was defined as the dominant uORF (#1). EC^2 represents the second highest uORF TE and the same goes for EC^3 and ">3". Each dot represents the median TE of a sample. (D) Fraction of conserved uORFs among dominant uORFs and other translated uORFs in each sample. The paired samples in Dm and Ds were linked together. The P value was obtained by the paired Wilcoxon signed rank test. ***, $P < 0.001$. (E) Absolute values of the interspecific TE fold changes ($\log_2 TE-FC$) of dominant uORFs and the other translated uORFs in each sample. The paired samples in Dm and Ds were linked together. The median value of each sample is shown. The P value was obtained via the paired Wilcoxon signed rank test. ***, $P < 0.001$. Data from the female head sample were used as an example in the right panel.

1,000 times (Fig. S14). This trend was also observed in *D. simulans* and persisted under different thresholds for defining “translated uORFs” (Fig. S14). Furthermore, the dominant uORFs showed a higher proportion of conserved uATGs than the other translated uORFs (median proportion 82.5% versus 78.8%, $P < 0.001$) (Fig. 2D). Additionally, TE fold-changes ($|\log_2 \text{TE-FC}|$) between the two species were, on an average, 23.2% smaller for dominant uORFs than for other conserved uORFs (Fig. 2E), suggesting that dominant uORFs are more likely under stronger stabilizing selection. These findings suggest that, in genes with multiple uORFs, the dominantly translated uORF may play a more important role in regulating CDS translation than the other uORFs.”

uORFs and CDSs show correlated translation differences between *D. melanogaster* and *D. simulans*

To investigate the relationship between translational changes in uORFs and their downstream CDSs across species, we analyzed the translation efficiency (TE) of uORFs and corresponding downstream CDSs in *D. melanogaster* and *D. simulans*. Consistent with our simulations that the translation of a uORF is tightly linked to that of downstream CDS, uORFs exhibited a significant positive correlation with the TE of their downstream CDSs in all samples analyzed ($P < 0.001$, Spearman’s correlation) (Fig. 3A). We then compared the interspecific TE change of a uORF ($\beta_u = \text{TE}_{u\text{ORF},\&.2}/\text{TE}_{u\text{ORF},2\text{CD}}$) with that of its corresponding CDS ($\beta_c = \text{TE}_{\text{CDS},\&.2}/\text{TE}_{\text{CDS},2\text{CD}}$) between *D. melanogaster* and *D. simulans*. We found β_u is significantly positively correlated with β_c across all samples (P values < 0.001 , Spearman’s correlation) (Fig. 3B). These results align well with our simulations, which showed that fluctuations in translational factors (such as ribosomes) influence both uORF and CDS translation in the same direction (Fig. S8-S9).

uORFs buffer interspecific translational divergence of CDSs

While the direction of translational efficiency (TE) changes for uORFs and CDSs tends to be consistent, our simulations suggest that the magnitude of TE changes in CDSs is generally smaller due to the buffering effect of uORF translation (Fig. 1B, S4-5). To validate this, we first identified uORFs and CDSs with significant interspecific TE differences by assessing whether β_u or β_c significantly deviated from 1, using an established statistical framework²³. This analysis uncovered 1,151 to 4,189 CDSs with significant interspecific TE changes (FDR < 0.05) (Table 1), with genes involved in development, morphogenesis, and differentiation being significantly enriched during embryonic stages (Fig. S15). Conversely, genes related to metabolism, response to stimuli, and signaling were enriched in larval, pupal, and adult stages (Fig. S15). Additionally, we identified 144 to 1,193 uORFs with significant TE differences between species, accounting for approximately 1-15% of expressed uORFs (Table 1). The smaller number of uORFs showing significant TE changes compared to CDSs between *D. melanogaster* and *D. simulans* likely reflects their shorter length and reduced statistical power, rather than indicating that uORFs are less variable in translation than CDSs.

To further investigate, we quantitatively compared the magnitude of interspecific TE changes between uORFs and their corresponding CDSs using a previous method²³. We defined $\gamma = \beta_c/\beta_u$ for a uORF-CDS pair within the same mRNA and tested whether γ was significantly different from 1 to identify pairs with differential TE changes between uORFs and CDSs (Fig. S16). When $\gamma < 1$ and $\beta_u > 1$, or $\gamma > 1$ and $\beta_u < 1$, it indicates that the TE change for the CDS is smaller than that for the uORF (Fig. S16). Among CDS-uORF pairs where $\beta_u > 1$, nearly all (8-487) showed a significant $\gamma < 1$ in each sample, except for one pair from the pupal stage where $\gamma > 1$ (Table 1). This suggests that the magnitude of TE changes in CDSs was generally smaller than in uORFs when uORF TE increased, and vice versa when uORF TE decreased ($\beta_u < 1$) (Table 1). These comparative translome analyses indicate that uORFs buffered mRNA translation changes during the evolutionary divergence of *D. melanogaster* and *D. simulans*.

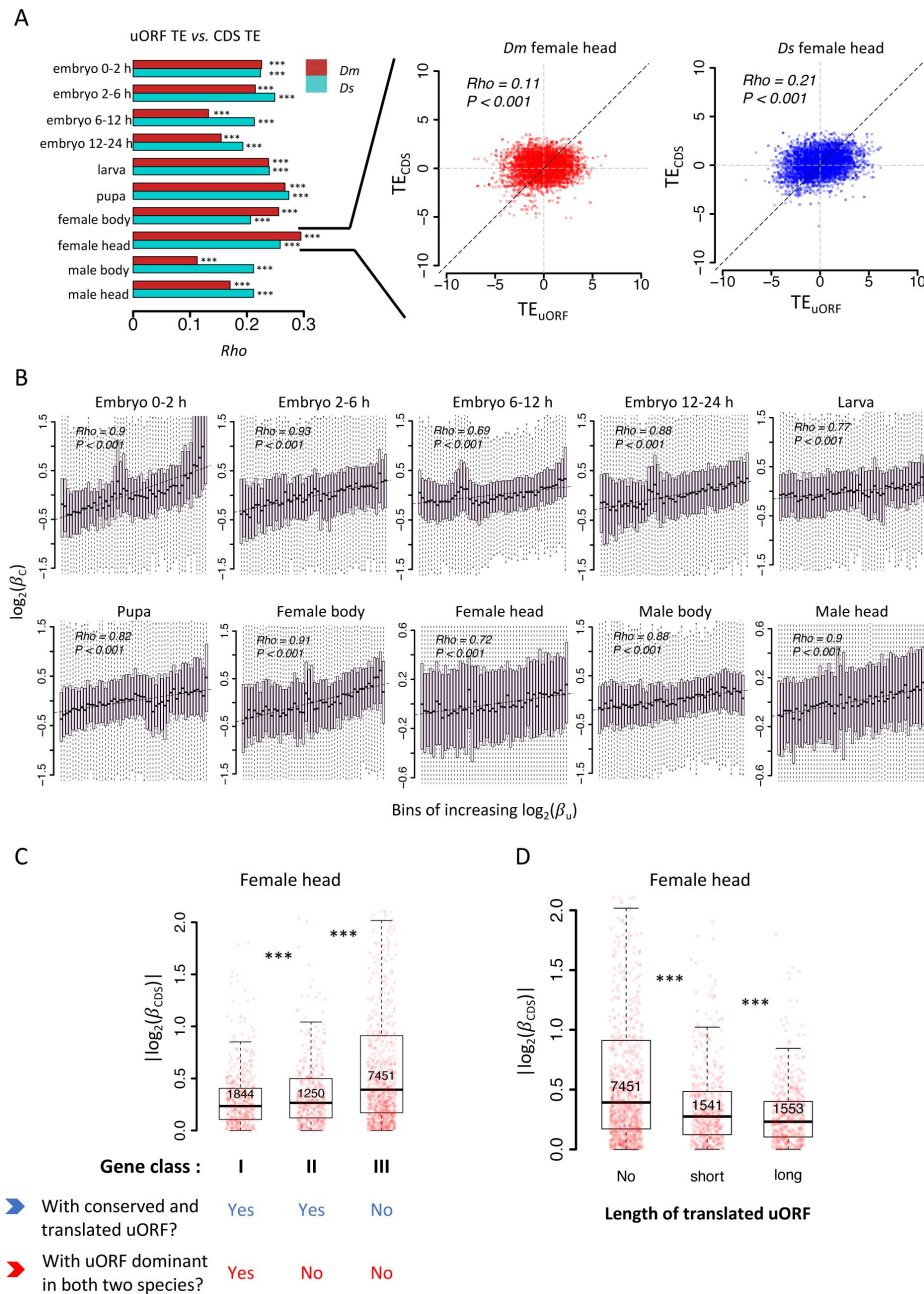


Figure 3.

uORFs reduce CDS translational divergence between *D. melanogaster* and *D. simulans*.

(A) The correlation of uORF TEs and the corresponding CDS TEs in 10 samples of *Dm* (*D. melanogaster*) and *Ds* (*D. simulans*). The bars represent Spearman's correlation coefficient (Rho). In all samples, we obtained both P values < 0.001 . Data for the female head sample of *Dm* and *Ds* are shown as examples in the right panel. (B) Correlations between interspecific uORF TE changes ($\log_2\beta_U$) and CDS TE changes ($\log_2\beta_C$) in 10 samples. The x-axis was divided into 50 equal bins with increasing β_U . Spearman's correlation coefficients (Rho) are shown at the top left. ***, $P < 0.001$ in the correlation test. (C) Genes expressed in female heads (mRNA RPKM > 0.1 in both species) were classified into three classes according to whether a gene had a conserved and dominantly translated uORF or not. Boxplots showing interspecific CDS TE variability $|\log_2(\beta_C)|$ of different gene classes. P values were calculated using Wilcoxon rank sum tests between the neighboring groups. ***, $P < 0.001$. (D) Genes expressed in female heads were classified into three classes according to the length of translated uORFs. Boxplots showing interspecific CDS TE variability $|\log_2(\beta_C)|$ of different gene classes. P values were calculated using Wilcoxon rank sum tests between the neighboring groups. ***, $P < 0.001$.

Sample types	# of expresse d uORFs *	$\beta_u \neq 1$ (%) **	# of expresse d CDSs *	$\beta_c \neq 1$ (%) **	uORF-CDS pairs with $\beta_u > 1$			uORF-CDS pairs with $\beta_u < 1$		
					Total	$\gamma > 1$	$\gamma < 1$	Total	$\gamma > 1$	$\gamma < 1$
0-2 h embryo	7,704	1,193 (15.49)	7,934	4,189 (52.80)	770	0	366	423	69	0
2-6 h embryo	7,822	567 (7.25)	9,795	3,063 (31.27)	249	0	63	318	135	0
6-12 h embryo	10,400	1,040 (10.00)	10,643	2,924 (27.47)	973	0	641	67	22	1
12-24 h embryo	11,365	535 (4.71)	11,158	3,537 (31.70)	454	0	234	81	8	2
Larva	10,008	464 (4.64)	11,831	3,554 (30.04)	110	0	26	354	177	0
Pupa	12,309	635 (5.16)	12,209	4,087 (33.48)	136	1	15	499	210	0
Male body	10,894	197 (1.81)	12,284	2,432 (19.80)	122	0	29	75	4	0
Male head	10,904	144 (1.32)	10,447	1,151 (11.02)	119	0	31	25	5	0
Female body	9,809	340 (3.47)	11,002	3,605 (32.77)	279	0	17	61	3	0
Female head	10,935	332 (3.04)	10,545	1,270 (12.04)	324	0	152	8	3	0

* Only uORFs and CDSs with an mRNA RPKM > 0.1 in both *D. melanogaster* and *D. simulans* were considered in each sample pair in the analysis.

** $\beta_u = TE_{uORF,sim}/TE_{uORF,mel}$ is the fold change of TE_{uORF} in *D. simulans* relative to *D. melanogaster* for each sample. $\beta_c = TE_{CDS,sim}/TE_{CDS,mel}$ is the fold change of TE_{CDS} in *D. simulans* relative to *D. melanogaster* for each sample. $\gamma = \beta_c/\beta_u$. For each CDS–uORF pair, either $\beta_u > 1$ and $\gamma < 1$ or $\beta_u < 1$ and $\gamma > 1$ means that the magnitude of TE change is lower for a CDS than for a uORF. The statistical significance of β_c , β_u , and γ were determined according to an FDR < 0.05.

Table 1.

Numbers of genes showing different magnitudes of TE changes between uORFs and CDS at the interspecific level.

uORF buffering is influenced by its conservation, dominance, and length

To investigate how the conservation level and translation patterns of uORFs influence their buffering capacity on CDS translation, we categorized genes expressed in each pair of samples into three classes: Class I, genes with uORFs conserved and dominantly translated in both *Drosophila* species; Class II, genes with conserved uORFs translated in both species but not dominantly in at least one; and Class III, the remaining expressed genes. We then compared the interspecific translation efficiency (TE) changes of the CDS ($|\beta_c|$) across these three categories. Significant differences in $|\beta_c|$ were observed, with a consistent hierarchy of Class I < II < III across all pairs of samples (Fig. 3C and S17). On average, Class I genes exhibited an average of 8.18% and 23.8% lower $|\beta_c|$ values compared to Class II and Class III, respectively. This indicates that conserved and dominantly translated uORFs exert a stronger buffering effect on CDS translation.

To further validate the simulation results suggesting that longer uORFs have a stronger buffering effect (Fig. 1C and S6-7), we divided genes expressed in each pair of samples into three groups: those without translated uORFs (No), those with short uORFs (short, total length below the median), and those with long uORFs (long, total length above the median). Consistently, longer uORFs were associated with stronger buffering effects on CDS translation across all pairs of samples (Fig. 3D and S18). Specifically, genes with longer uORFs showed 12.7% and 26.5% lower $|\beta_c|$ values compared to genes with short uORFs or no uORFs, respectively.

Overall, these findings underscore that the buffering capability of a uORF is positively correlated with its conservation level, translation dominance, and length.

uORFs buffer translational fluctuations during *Drosophila* development

Gene expression undergoes dynamic changes during *Drosophila* development, with significant alterations in the translation program to meet developmental demands, including shifts in ribosome loading rates^{24,62,63}. Therefore, we extended our analysis to investigate the role of uORFs in buffering these translational fluctuations, hypothesizing that uORFs could mitigate them during development. According to our hypothesis, if a gene has a translated uORF in *D. melanogaster* but not in its orthologous gene in *D. simulans*, then the translation of this gene is likely more stable across developmental stages in *D. melanogaster* than its ortholog in *D. simulans*, and vice versa.

To test this hypothesis, we compared the CV of CDS TE across 10 developmental stages in *D. melanogaster* and in *D. simulans*, respectively. Genes with translated uORFs (TE > 0.1 in at least one sample) in *D. melanogaster* exhibited significantly 22.5% smaller CVs in *D. melanogaster* than their orthologs lacking these uORFs in *D. simulans*, indicating more stable translation (Fig. 4A). Consistently, for genes with translated uORFs in *D. simulans* but not in *D. melanogaster*, the CVs were 13.3% lower in *D. simulans* compared to their orthologs lacking these uORFs in *D. melanogaster* (Fig. 4B). Consistent results were observed when a uORF is required to be translated (TE > 0.1) in all 10 samples (Fig. S19A), in the 4 embryonic stages (Fig. S19B) or in the 6 stages including embryos, larva, and pupa (Fig. S19C). Moreover, within each species, genes with translated uORFs also showed less variability in CDS TE across developmental stages compared to those without translated uORFs, with a 31.8% reduction in *D. melanogaster* and a 28.9% reduction in *D. simulans* (Fig. 4C). This effect was consistent across different thresholds for defining “translated uORFs” (Fig. S20).

These results suggest that uORFs function as translational buffers, reducing gene translation fluctuations during *Drosophila* development.

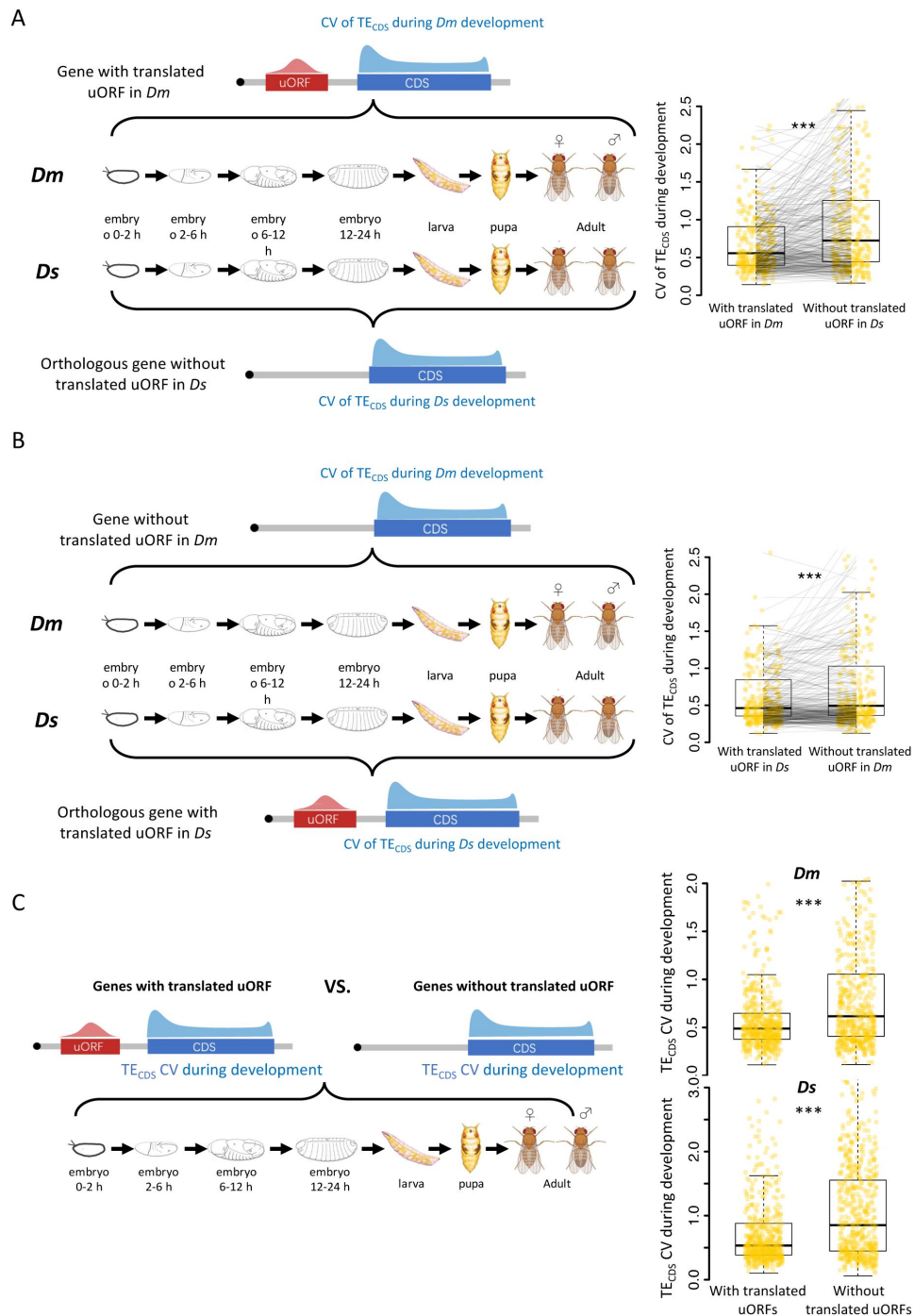


Figure 4.

uORFs could reduce CDS translational fluctuation during *Drosophila* development.

(A) The CV of TE_{CDS} across 10 *Dm* (*D. melanogaster*) samples and 10 *Ds* (*D. simulans*) samples. The selected gene with uORFs translated ($TE > 0.1$) in at least one *Dm* sample but its homologous gene without translated uORF in *Ds* samples. Each pair of dots linked by a gray line represents a pair of homologous genes in *Dm* and *Ds*. ***, $P < 0.001$, Wilcoxon signed-rank test. (B) The CV of TE_{CDS} across 10 *Dm* samples and 10 *Ds* samples. The selected gene with uORFs translated ($TE > 0.1$) in at least one *Ds* sample but its homologous gene without translated uORF in *Dm* samples. Each pair of dots linked by a gray line represents a pair of homologous genes in *Dm* and *Ds*. ***, $P < 0.001$, Wilcoxon signed-rank test. (C) Within each *Drosophila* species, the CV of TE_{CDS} of genes with translated uORFs compared to genes without the translated uORFs. The P values are obtained by the Wilcoxon rank sum test. ***, $P < 0.001$.

Knocking out the uORF of *bcd* increased *bcd* CDS translation in *D. melanogaster*

After verifying the uORFs-mediated translational buffering during *Drosophila* evolution and development, we next aimed to directly explore the biological function of these buffering-capable uORFs in vivo. We first applied stringent criteria to identify uORFs with significant buffering effects on CDS translation between *D. melanogaster* and *D. simulans*. Specifically, we looked for 1) uORFs with significant TE changes ($|\log_2(\beta_u)| > 1.5$, adjusted $P < 0.05$), 2) negligible changes in its corresponding CDS translation ($|\log_2(\beta_u)| < 0.05$, adjusted $P > 0.05$), and 3) a significant difference between the magnitude of these changes ($|\log_2(y)| > 1.5$, adjusted $P < 0.05$). We identified 131 uORF-CDS pairs in 103 genes that meet these criteria in at least one stage/tissue (Table S3), with a majority of the genes (67%, 69 out of 103) from embryonic stages (Fig. S21A), suggesting a crucial role for uORFs in maintaining translational stability during early development. Among these genes, one notable case is the *bicoid* (*bcd*) gene, a master regulator of anterior-posterior axis patterning during early embryogenesis^{64–66}. The *bicoid* gene contains a 4-codon uORF (excluding the stop codon) in its 5' UTR, and branch length score (BLS) analysis¹⁸ showed that the start codon (uATG) of the *bicoid* uORF is highly conserved across the *Drosophila* phylogeny, with a BLS of 0.90 (on a scale from 0 to 1, where higher values indicate greater conservation) (Fig. 5A and S21B). Ribo-Seq data revealed that in 0–2 h embryos, the TE of the *bcd* uORF varied more than threefold, while the TE of the CDS was virtually the same between *D. melanogaster* and *D. simulans* (Fig. 5B and Table S3). This suggests that the *bcd* uORF buffers translation during early development.

To investigate the regulatory role of the *bcd* uORF, we used CRISPR-Cas9 to knock out its start codon in *D. melanogaster*, generating two mutant homozygotes (uKO1/uKO1 and uKO2/uKO2) with a genetic background matched to that of the wild-type (WT) (Figs. 6A and S22). To determine whether these mutations enhanced *bcd* CDS translation, we performed ribosome fractionation followed by qPCR^{67–69}, comparing *bcd* mRNA levels in polysome and monosome fractions (P-to-M ratio) from 0–2 h embryos of uKO1/uKO1, uKO2/uKO2, and WT flies (Fig. 6B). A larger P-to-M ratio means more mRNAs are enriched in the polysome fractions and bound by more ribosomes, thus indicative of higher translation efficiency. P-to-M ratios were higher in the mutants compared to WT at 29°C (Fig. 6C), with a similar but less pronounced trend at 25°C, where the difference between uKO2/uKO2 and WT was not statistically significant (Fig. 6C). These findings, along with the known impact of temperature on gene expression and phenotypic plasticity^{70,71}, suggest that the regulatory function of the uORF and overall translation efficiency are temperature-sensitive, highlighting a complex interplay between environmental conditions and gene regulation. To further confirm the *bcd* uORF's regulatory function, we conducted dual-luciferase reporter assays. The 5'UTR from *bcd* mutant (uKO1 or uKO2) or WT was cloned into a *Renilla* luciferase reporter construct (Fig. 6D). Luciferase activity was significantly higher in uKO1 and uKO2 compared to the WT 5'UTR, confirming the repressive role of the *bcd* uORF.

bcd uORF mutants show wide transcriptomic alteration during *Drosophila* embryogenesis

Since Bcd regulates the expression of many zygotic genes^{64–66}, we anticipated that the increased translation of *bcd* resulting from disrupting the uORF would influence *Drosophila* transcriptomes and phenotypes. To verify this notion, we performed RNA sequencing on embryos from WT and uKO2/uKO2 flies at four developmental stages (0–2 h, 2–6 h, 6–12 h, and 12–24 h) under both 25°C and 29°C conditions, using two biological replicates (Table S4 and Fig. S23). Differential expression analysis revealed widespread alterations in gene expression between WT and mutant embryos, with the number of differentially expressed genes (DEGs) increasing over

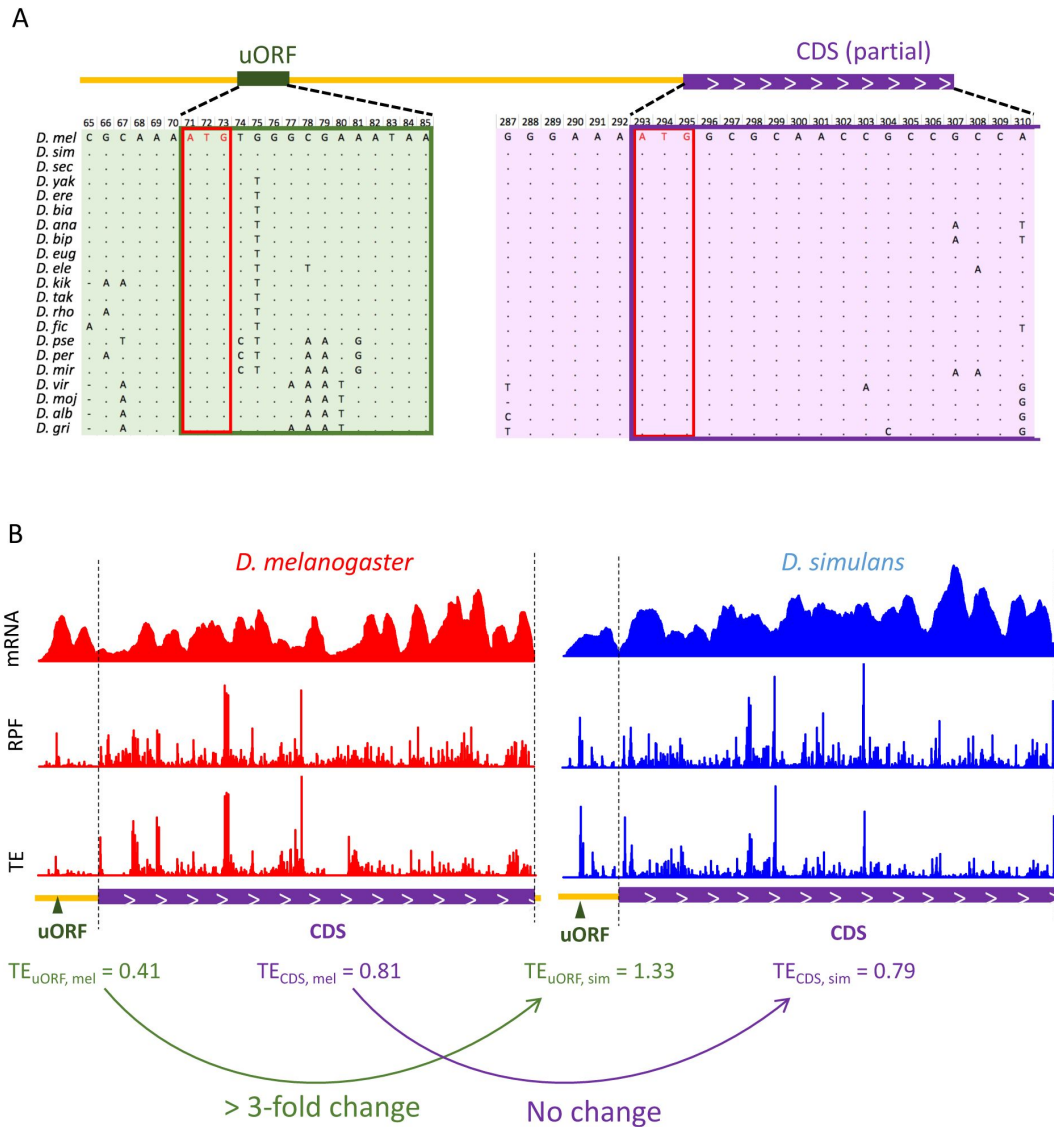


Figure 5.

The strong buffering uORF of *bcd* and its knockout.

(A) Multiple sequence alignment of the *bcd* uORF and partial CDS in *D. melanogaster* and 20 other *Drosophila* species. The uORF and CDS are boxed in green and purple, respectively. The start codons of the uORF and CDS are boxed in red. (B) The coverage of mRNA-Seq (top), Ribo-Seq (middle), and TEs (bottom) of the *bcd* uORF and CDS in 0-2 h embryos of *D. melanogaster* (red) and *D. simulans* (blue). The uORF and CDS are denoted at the lower panel with dark green triangles and purple boxes, respectively. The 2 dashed lines mark the CDS region. The uORF TE, CDS TE and their interspecific changes were labeled at the bottom. (C) Genotypes of WT and two uORF knock-out strains (uKO1 and uKO2) generated by CRISPR-Cas9 technology. The uORF is boxed in dark green, and the red ATG represents the start codon of the uORF in the *D. melanogaster* genome.

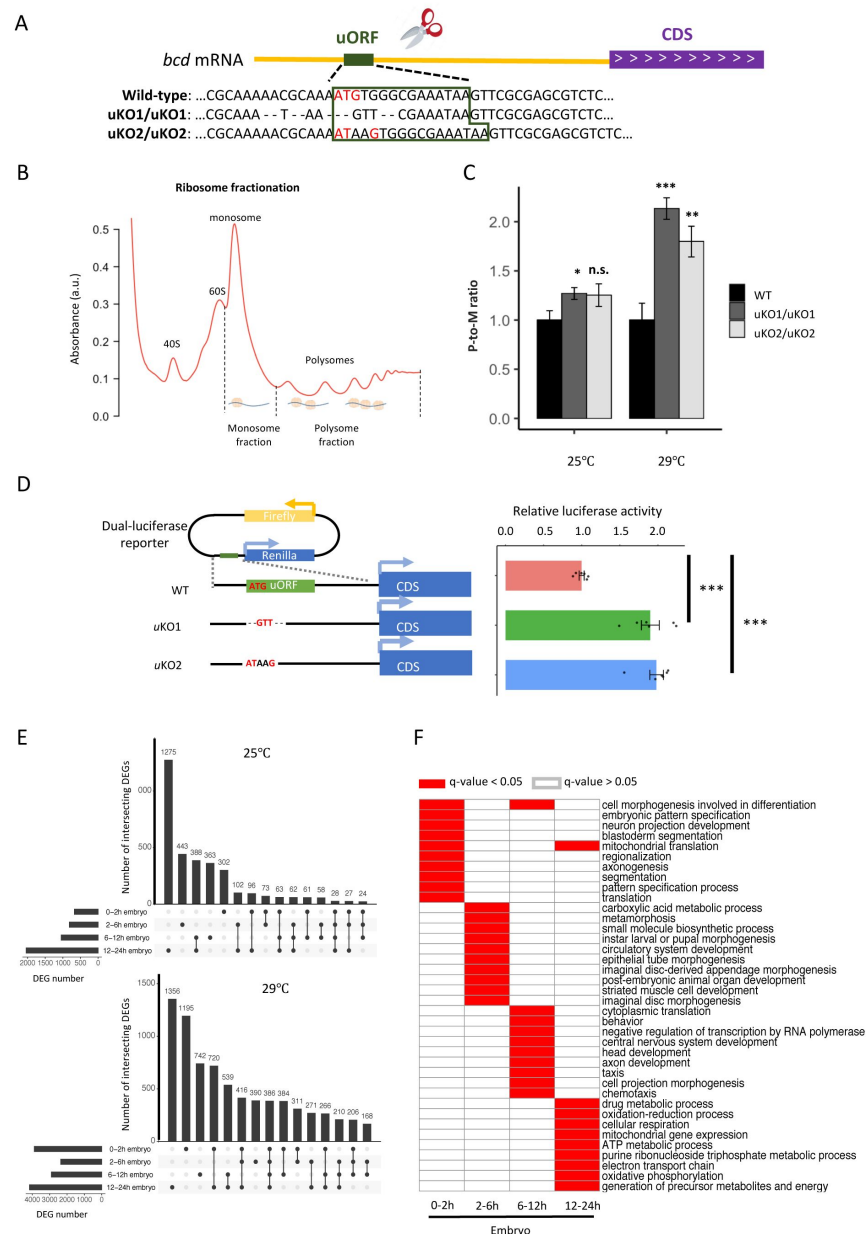


Figure 6.

Knocking out the *bcd* uORF increases CDS translation and perturbs the transcriptome during *D. melanogaster* embryogenesis.

(A) Dual-luciferase assay for *bcd* WT uORF and mutated uORF. The reporter structures of the WT and uORF mutants are illustrated on the left. The uORF mutant sequence was the same as that in the fly mutant created with CRISPR-Cas9 technology. The relative activity of *Renilla* luciferase was normalized to that of firefly luciferase. Error bars represent the S.E. of six biological replicates. Asterisks indicate statistical significance (***, $P < 0.001$). (B) Two ribosome fractions (monosome and polysome) of 0-2 h embryos were separated in a sucrose density gradient. Relative RNA abundance in the monosome and polysome fractions was quantified by real-time quantitative PCR. (C) P-to-M ratio of *bcd* mRNA (*bcd* mRNA abundance in polysome fraction/*bcd* mRNA abundance in monosome fraction) at 25°C (left) and 29°C (right). The P-to-M ratios of mutants were normalized to WT controls at 25°C and at 29°C, respectively. Error bars represent the S.E. of six biological replicates. Asterisks indicate statistical significance (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; n.s., $P > 0.05$). (D) The number of DEGs in each stage and their intersection with each other at 25°C (top) and 29°C (bottom). (E) Gene ontology analysis of DEGs at 29°C in each stage. The biological process (BP) terms with q-values < 0.05 in each stage are indicated in red and others are indicated in white.

developmental time and at higher temperature. At 25°C, we identified 674, 817, 1047, and 2,041 DEGs in 0-2 h, 2-6 h, 6-12 h, and 12-24 h embryos, respectively; while at 29°C, we detected 3,884, 2,358, 2,901, and 4,164 DEGs in the corresponding stages (**Fig. 6D**). The majority of DEGs were stage-specific, with only a small fraction consistently differentially expressed across all four stages. Functional enrichment analysis of the DEGs revealed distinct biological pathways affected at each stage, including cell morphogenesis and pattern specification in 0-2 h embryos, metabolic processes and tissue development in 2-6 h and 6-12 h embryos, and mitochondrial respiration in 12-24 h embryos (**Fig. 6E**). Notably, direct targets of Bcd ⁷² were significantly enriched among the DEGs in three out of the four stages, with the exception of 2-6 h embryos (Fig. S24). RT-qPCR validation of 20 target genes of Bcd confirmed the reliability of the RNA-seq differential expression analysis (Fig. S25). Together, these findings demonstrate that disruption of the *bcd* uORF leads to widespread transcriptional changes during *Drosophila* development, affecting processes ranging from embryogenesis to postembryonic metabolism.

***bcd* uORF mutants display decreased hatching rates and starvation resistance**

Given the widespread transcriptome alterations, we anticipated phenotypic abnormalities in the *bcd* uORF mutants. As expected, both uKO1/uKO1 and uKO2/uKO2 mutants exhibited significantly lower hatching rates compared to WT ($P = 1.4 \times 10^{-6}$ and 7.9×10^{-6} , respectively, Wilcoxon rank sum test [WRST]; **Fig. 7A**). At 25°C, uKO1/uKO1 mutants produced fewer offspring than WT flies ($P < 0.001$, WRST, **Fig. 7B**). Given that *bcd* is a maternal gene, we expected reciprocal crosses between uKO1/uKO1 mutants and WT flies to produce different outcomes. Indeed, crossing uKO1/uKO1 males with WT females resulted in offspring numbers similar to those from WT crosses, while crossing uKO1/uKO1 females with WT males yielded offspring numbers comparable to crosses between uKO1/uKO1 mutants (**Fig. 7B**). This verified that the reduction in offspring is due to maternal defects in the uKO1/uKO1 mutants. Similar patterns were observed for uKO2/uKO2 mutants at 25°C (**Fig. 7B**). Notably, the fecundity reduction in uORF-KO mutants was more pronounced at 29°C (**Fig. 7C**). Furthermore, crossing uKO1/uKO1 and uKO2/uKO2 mutants produced significantly fewer progeny than WT flies (**Fig. 7C**), ruling out genetic background or off-target effects. Collectively, these data demonstrate that disrupting the *bcd* uORF significantly impaired hatchability and fertility.

We also found both uKO1/uKO1 and uKO2/uKO2 female mutants perished significantly faster than WT flies under starvation conditions (**Fig. 7D**). Males showed similar tendencies, although the difference was not statistically significant for uKO1/uKO1 mutants (**Fig. 7D**). These data suggest that the knockout of *bcd* uORF diminished starvation resistance in adults, likely due to embryogenesis abnormalities induced by the *bcd* uORF deletion, even in those that successfully developed to adulthood.

Conservation of uORF-mediated translational buffering in primates

To explore the generality of uORF-mediated translational buffering across evolutionary clades, we analyzed previously published transcriptome and translome data from three tissues (brain, liver, and testis) in humans and macaques ⁷³. We identified 33,680 canonical uORFs in humans and 29,516 in macaques, with 24,385 conserved between the two species. Despite the larger number of uORFs in primates compared to *Drosophila* due to differences in genome size and gene number, the median TE of conserved uORFs was 1.79 times that of non-conserved uORFs in humans, and the corresponding ratio was 3.43 in macaques (**Fig. 8A** and Fig. S26). TEs of uORFs were positively correlated between humans and macaques across all tissues ($P < 0.001$, **Fig. 8B**). Additionally, significant positive correlations were observed between the TEs of uORFs and their corresponding coding sequences (CDSs) in all tissues (Fig. S27). Although interspecific TE divergence of uORFs (β_u) and CDSs (β_c) were positively correlated ($P < 0.001$, **Fig. 8C**), uORFs

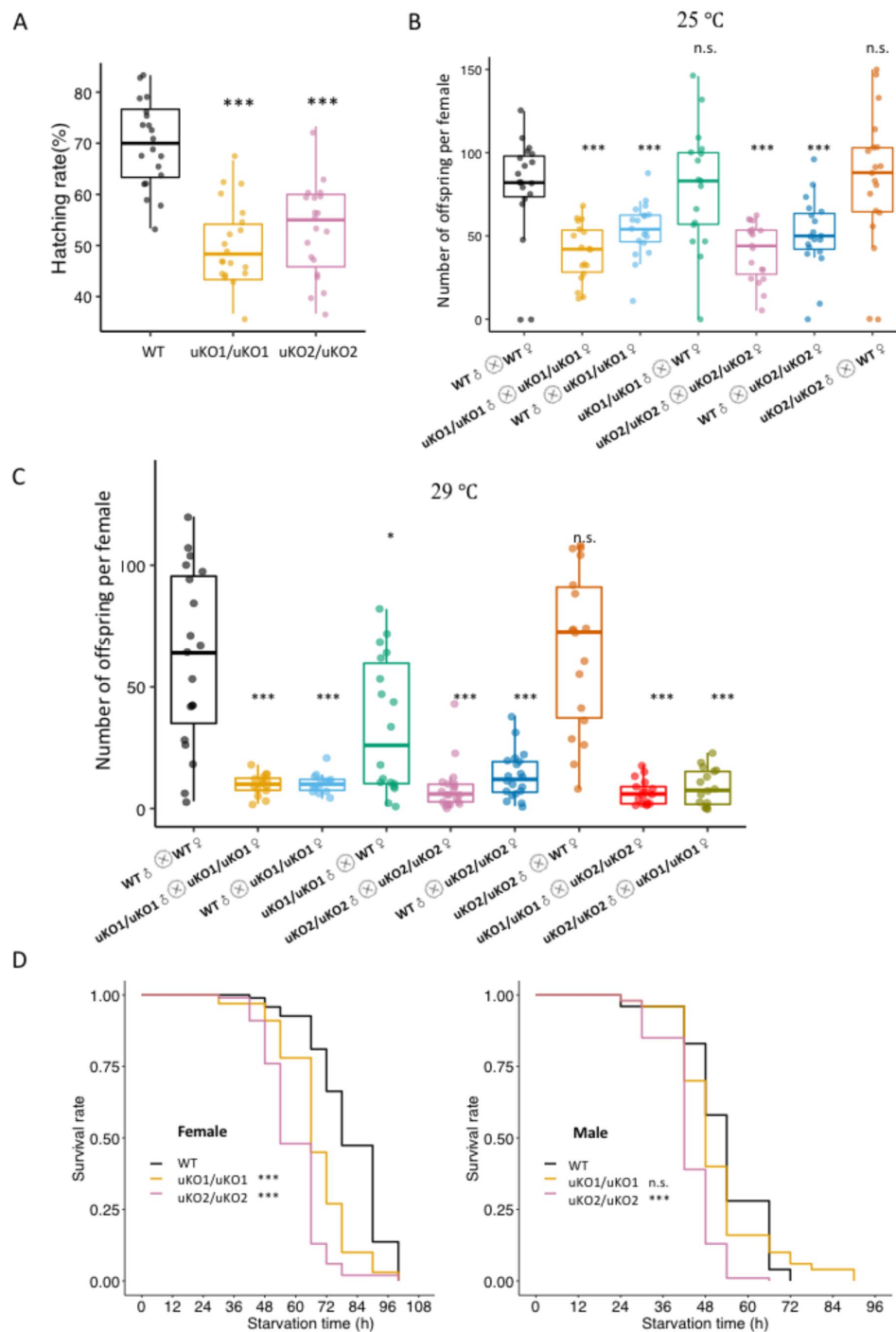


Figure 7.

Knockout of the *bcd* uORF reduces offspring number and starvation resistance.

(A) Comparison of the hatching rates (%) of mutant and WT offspring (n=20, Wilcoxon rank sum test; ***, $P < 0.001$). (B) The offspring number per maternal parent in different crosses over 10 days at 25°C. Asterisks indicate significant differences between various crosses and crosses of WT females with WT males (n=20, Wilcoxon rank sum test; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; n.s., $P > 0.05$). The different crosses were denoted as the x-axis labels. (C) The offspring number per maternal parent in different crosses over 10 days at 29°C. (D) Survival curves of WT and mutant adult flies of females (left) and males (right) under starvation conditions. The black line represents the WT, the red line represents the uKO1/uKO1 mutant, and the blue line represents the uKO2/uKO2 mutant. Asterisks indicate significant differences compared to the WT. (n=200, log-rank test; ***, $P < 0.001$; n.s., $P > 0.05$).

generally exhibited larger divergence (Table S5). Notably, longer uORFs showed stronger buffering effects on CDS translation, reducing interspecific TE divergence by 12.9% and 38.0% compared to genes with short or no uORFs (**Fig. 8D** and Fig. S28).

As in *Drosophila*, we categorized expressed human genes into three classes based on uORF conservation and translation: Class I, genes with conserved and dominantly translated uORFs in both humans and macaques; Class II, genes with conserved uORFs translated in both species but not dominantly in at least one; and Class III, the remaining expressed genes (**Fig. 8E** and Fig. S29). Consistent with findings in *Drosophila*, significant differences in $|\beta_c|$ between humans and macaques were observed in the order of Class I < II < III, with Class I genes showing 9.8% and 17.1% lower $|\beta_c|$ values compared to Class II and Class III genes, respectively (**Fig. 8E** and Fig. S29). These similarities between primates and *Drosophila*—two clades that diverged over 700 million years ago⁷⁴—suggest that uORF-mediated translational buffering is a widespread mechanism for stabilizing gene translation across evolutionary clades.

We also analyzed matched mRNA-Seq and Ribo-Seq data from 69 human lymphoblastoid cell lines^{75,76} to test whether uORFs buffer against translational variability across different individuals within humans. Genes with translated uORFs exhibited, on average, 8.65% lower CVs in CDS TE across individuals than genes without translated uORFs (**Fig. 8F**), with longer uORFs showing stronger buffering effects (Fig. S30). Collectively, these findings suggest that uORFs play a crucial role in reducing translational variability both across evolutionary clades and within species.

Discussion

Translational control is vital for maintaining protein homeostasis and cellular activities^{77,78}. However, the mechanisms underlying the conservation of protein abundance across species remain largely unknown^{5,6}. In this study, we extended the ICIER model and conducted simulations to explore uORFs' regulatory roles in buffering translation during evolution. Our simulations demonstrated that uORF translation reduces variability in downstream CDS translation, with buffering capacity positively correlated with uORF translation efficiency, length, and number. Comparative transcriptome analyses across developmental stages of two *Drosophila* species provided evidence that uORFs mitigate interspecific differences in CDS translation. Similar patterns were observed between humans and macaques, and between human individuals, suggesting that uORF-mediated buffering is an evolutionarily conserved mechanism in animals. Additionally, *in vivo* experiments showed that knocking out the *bcd* uORF led to aberrant embryogenesis and altered starvation resistance, with more pronounced phenotypic effects at higher temperatures, supporting the role of uORFs in fine-tuning translation and phenotypic outcomes.

While the prevailing consensus on uORFs' functions is that they repress the translation of downstream CDS by sequestering ribosomes^{34,43–48}, recent studies based on suggest uORFs might play essential roles in buffering translational noise and stabilizing protein expression^{43,55,56}. While these studies have primarily focused on the role of uORFs in single-gene cases or under stress conditions, our work expands this knowledge by demonstrating that uORFs act as a general mechanism for buffering translational variability on a genome-wide scale across multiple species and developmental stages. Our findings also suggest that uORF-mediated translational buffering is not merely a response to environmental stress but an evolutionarily conserved mechanism integral to maintaining protein homeostasis across species. Organisms can buffer phenotypic variation against environmental or genotypic perturbations through “canalization”⁷⁹. The discovery that uORFs are crucial for *Drosophila* development links them to canalization, providing evidence that uORFs reduce interspecific CDS translation differences and enhance phenotypic stability. While previous studies primarily focused on the stabilization of transcriptional level^{80–84} or protein level^{55,82,85–87}, our study opens new avenues

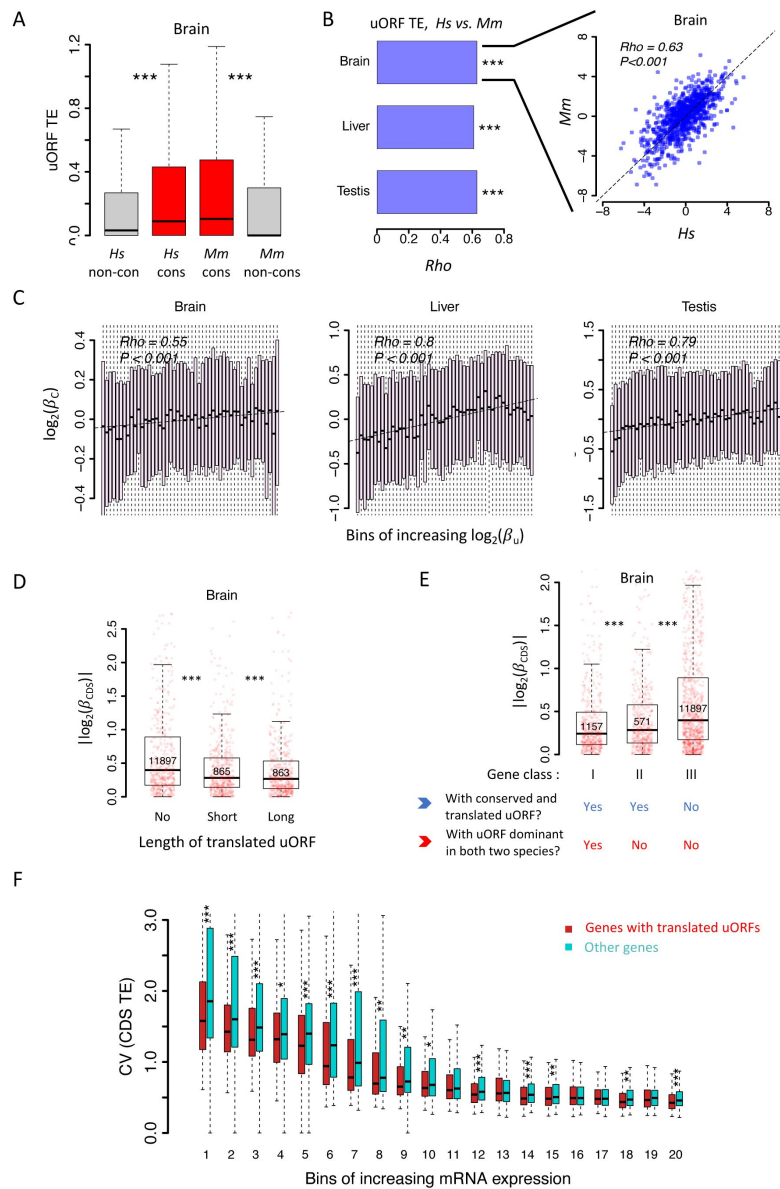


Figure 8.

uORFs function as translational buffers in primates.

(A) Boxplots showing the TEs of conserved and species-specific uORFs between *Hs* (*H. sapiens*) and *Mm* (*M. mulatta*). Data for the brain is shown as an example. Wilcoxon rank sum tests. ***, $P < 0.001$. (B) Spearman's correlation coefficient (Rho) of uORFs' TE between humans and macaques. The Rho values in the brain, liver, and testis were shown as bar plots. ***, $P < 0.001$. Data for the brain is shown as an example in the right panel. (C) Correlation between interspecific uORF TE changes ($\log_2\beta_u$) and corresponding CDS TE changes ($\log_2\beta_c$) in three tissues. The x-axis was divided into 50 equal bins with increasing β_u . (D) Genes expressed in brains were classified into three classes according to the total length of translated uORFs. Boxplots showing interspecific CDS TE variability $|\log_2(\beta_c)|$ of different gene classes. P values were calculated using Wilcoxon rank sum tests between the neighboring groups. ***, $P < 0.001$. (E) Genes expressed in brains (mRNA RPKM > 0.1 in both species) were classified into three classes according to whether a gene had a conserved and dominantly translated uORF (TE > 0.1) in both species or not. Boxplots showing interspecific CDS TE variability $|\log_2(\beta_c)|$ of different gene classes. P values were calculated using Wilcoxon rank sum tests between the neighboring groups. ***, $P < 0.001$. (F) Boxplot showing the coefficients of variation (CVs) of CDS TE among the 69 lymphoblastoid cell lines (LCLs). Expressed genes (mean mRNA RPKM > 0.1) were divided into 20 bins with increased mRNA expression levels. In each bin, the genes were divided into two fractions according to whether the gene had a translated uORF or not. Wilcoxon rank sum tests. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

for exploring how organisms maintain phenotypic robustness at the translational level through molecular mechanisms like uORF-mediated buffering. This study lays the groundwork for future research into the evolutionary and functional roles of uORFs, particularly in the context of canalization and adaptive evolution.

The original ICIER model was devised to elucidate how a single uORF can confer resistance to global translation inhibition on its host gene during stress conditions ⁴³. It proposed that an elongating ribosome (80S) causes downstream 40S subunits to dissociate from the mRNA after collision. Our revised model expands this scope to include bidirectional dissociation, where an 80S ribosome can release both upstream and downstream 40S subunits, reflecting recent findings that 80S primarily causes upstream 40S dissociation during 40S/80S collisions ⁵⁶. Mechanistically, increasing the loading rate of 40S ribosomes onto the mRNA (R_{in}) elevates the density of 40S and 80S ribosomes scanning the uORF, thereby raising the likelihood of 40S-80S collisions and subsequent 40S dissociation. This leads to a reduced flow of 40S ribosomes reaching the CDS relative to the initial increase in 40S at the 5' end of the mRNA and uORF. Conversely, a decrease in 40S ribosome loading rate reduces the density of scanning ribosomes, diminishing the probability of 40S-80S collisions and 40S dissociation, and partially restoring the flow of 40S ribosomes to the CDS. Overall, the 40S-80S collision mechanism and subsequent 40S dissociation at the uORF moderate downstream CDS translation less than changes in ribosome loading at the 5' end of the mRNA or uORF. The revised model underscores the analogy of uORFs as “molecular dams” adeptly modulating the stream of ribosomes and cushioning downstream CDS translation against fluctuations. Further, the revised model integrates ribosomal collisions and separations within the CDS located downstream of the uORF, shifting the focus from merely tracking 40S escape rates from the uORF to directly quantifying completed translations by 80S ribosomes at the terminus of both uORFs and CDSs. This methodology provides a more direct and quantitative assessment of CDS translation efficiency, offering new insights into the buffering role of uORFs. Additionally, our study broadens its perspective to account for the potential interactive effects of multiple uORFs within a gene, recognizing the prevalent clustering of these regulatory elements and their combinatorial influence on translational control. In essence, our study presents a nuanced and comprehensive expansion of the ICIER model, paving the way for a deeper understanding of uORFs as pivotal elements in the maintenance of protein synthesis stability under variable translational conditions. However, we realize the existence of alternative mechanisms governing uORF function, such as those found on traditional repression models ⁵⁵ or ribosome queuing and re-initiation strategies ⁵⁶. As such, further computational studies are needed to dissect and understand the full spectrum of uORF-mediated translational regulation.

The *bcd* gene, crucial for *Drosophila* embryogenesis, has been well studied ^{65,66,88–92}, but its uORF remains underexplored. Our study showed that deleting the *bcd* uORF led to abnormal phenotypes and reduced fitness, underscoring the importance of uORF-mediated translation. These effects were more severe under heat stress (29°C) than at normal temperatures (25°C), suggesting uORFs play a vital role in buffering phenotypic variation. Rescue experiments, typically used to confirm that phenotypic changes are due to specific genetic edits, faced challenges in our study and previous uORF-KO experiments in animals ^{93,94} and plants ^{95–98}. We altered the start codon of the *bcd* uORF to minimize effects on the 5' UTR, without impacting the coding sequence, making traditional rescue experiments impractical. To control for genetic background variations, we backcrossed two *bcd* uORF-KO mutant strains with w^{1118} flies for nine generations, using w^{1118} as a control. Both uORF-KO mutants (uKO1/uKO1 and uKO2/uKO2) consistently showed reduced progeny, ruling out background or off-target effects. Crosses between the two mutants produced fewer offspring than WT crosses, confirming that the reduced progeny was due to the uORF deletions, not genetic background. As *bcd* is maternally inherited, we expected defective phenotypes only in offspring from mutant females crossed with wild-type males. This was confirmed: uKO1/uKO1 males crossed with wild-type females had normal offspring, while uKO1/uKO1 females crossed with wild-type males had reduced offspring, similar to mutant crosses. Similar results were observed for uKO2/uKO2 mutants. These findings confirm that

disrupting *bcd* uORFs significantly reduces hatchability and fertility, effects not attributable to genetic background. Our results highlight the critical role of *bcd* uORFs in development and fecundity, and provide a basis for future studies on uORF regulation in other genes across species.

Taken together, this study demonstrates the role of uORF-mediated translational buffering in mitigating variability in gene translation during species divergence and development, using large-scale comparative transcriptome and translome data. Our work addresses gaps in understanding the stabilization of gene expression at the translational level and offers new insights into the evolutionary and functional significance of uORFs.

Materials and Methods

Modeling the uORF-mediated buffering effect on CDS translation

We adapted the stochastic ICIER framework developed by Andreev *et al.* [43](#) based on the TASEP model, with several major modifications. First, we concatenated a 500-codon CDS downstream of the uORF. A leaky scanning ribosome from the uORF would initiate translation at the CDS start codon with a probability of I_{CDS} . The probability of elongating ribosomes moving along the CDS ($v_{EC} = 0.5$) in each action was higher than that along uORFs ($v_{Eu} = 0.3$) in the model settings, considering that uORFs usually encode blocking peptides [99](#)–[103](#) or contain stalling codons [56](#), [104](#), [105](#). Second, we recorded the number of elongating ribosomes that completed translation at the stop codon of a CDS (N_{EC}) or uORF (N_{Eu}) during a given time period and regarded them as proxies for quantifying the CDS or uORF TE. Third, we considered three models for simulating the consequences of scanning ribosomes colliding with elongating ribosomes mentioned above: a downstream dissociation model, an upstream dissociation model, and a double dissociation model.

In detail, the mRNA molecule structure in the simulations consisted of five parts: a 5'-leader before the uORF (150 nucleotides), the uORF (ranging from 2 to 100 codons), a segment between the uORF and CDS (150 nucleotides), the CDS (500 codons) and the 3'UTR (150 nucleotides). The R_{in} determined the loading rate of the 40S scanning ribosomes on the mRNA 5'-terminus, and it roughly represented the availability of *trans* translational resources in the cell. We generated two sets of R_{in} values (Fig. S1A) to simulate the variation in the availability of translational resources (40S scanning ribosomes, etc.): i) 1000 values of R_{in} were generated from a random generator of a uniform distribution, $U(0, 0.1)$; ii) 1000 values were first generated from a random generator of an exponential distribution, $E(1)$, and then each of the 1000 values was divided by 70 to obtain 1000 values of R_{in} . I_{uORF} and I_{CDS} represent the probability of translation at the start codon of a uORF or CDS, respectively. We used different combinations of I_{uORF} and I_{CDS} to test how translational initiation strength influenced the buffering effect of uORFs. v_S determined the scanning rate of the 40S ribosome, and v_{Eu} and v_{EC} determined the elongation rate of the 80S ribosome at a uORF or CDS, respectively. All the parameters used in our simulation are listed in Table S1.

The mRNA molecule was modeled as an array, and the value of each position in the array represented the occupation status of ribosomes along the mRNA molecule. The process of translation was simulated by a series of discrete actions referred to Andreev *et al.* [43](#). Upon each action, there was a probability that certain events would occur, including the addition of a 40S ribosome to the 5'-terminus, the transformation of a 40S ribosome into an 80S ribosome at a CDS or uORF start codon, the movement of a 40S ribosome or 80S ribosome to the next codon in the CDS or uORF, etc. More specific operations in each action were described in Supplementary Materials and Methods. The simulation code is freely available at GitHub: https://github.com/lujlab/uORF_buffer.

After 1,000,000 actions, we recorded the number of 80S ribosomes that completed translation at the uORF (N_{Eu}) and CDS (N_{EC}) for each R_{in} input, which was used to represent the protein production rates (i.e. translation efficiency) of the uORF and CDS, respectively. To test the effects of different factors, we used various combinations of the parameters in Table S1 in the simulations. For a given distribution of R_{in} input (1000 values) and the dissociation model, we obtained the corresponding 1000 N_{EC} values and their CV values by calculating the ratio of the standard deviation to the mean of N_{EC} .

In the double-uORF simulation, we only adopted a uniform distribution of R_{in} input and the downstream dissociation model. Most parameters and simulation processes were similar to those in the single-uORF simulation, except that an additional uORF was introduced upstream of the CDS with an initiation probability of I_{uORF2} .

Annotation of uORFs in *D. melanogaster* and *D. simulans*

We downloaded the whole-genome sequence alignment (maf) of *D. melanogaster* (dm6) and 26 other insect species from the University of California Santa Cruz (UCSC) genome browser (genome.ucsc.edu)¹⁰⁶. To improve the genome correspondence between *D. melanogaster* and *D. simulans*, we adopted a newly published reference-quality genome of *D. simulans* (NCBI, ASM438218v1)¹⁰⁷ to replace the original *D. simulans* genome in UCSC maf. We softly masked repetitive sequences in the new genome by RepeatMasker 4.1.1 (<http://www.repeatmasker.org>)¹⁰⁸ and aligned it to dm6 with lastz¹⁰⁹ in runLastzChain.sh following UCSC guidelines. The lastz alignment parameters and the scoring matrix were the same as the parameters of dm6 and droSim1 in UCSC. Chained alignments were processed into nets by the chainNet and netSyntenic programs. The alignment was integrated into the multiple alignments of 27 species with multiz¹⁰⁹.

Based on the genome annotation of *D. melanogaster* (FlyBase r6.04, <https://flybase.org/>), we used the Galaxy platform to parse the multiple sequence alignments of 5'UTRs in *Drosophila*. The 5'UTR sequences of each annotated transcript from *D. melanogaster* and the corresponding sequences in *D. simulans* were extracted from the maf. The start codons of putative uORFs were identified by scanning all the ATG triplets (uATGs) within the 5'UTRs of *D. melanogaster* and *D. simulans*. uATGs that overlapped with any annotated CDS region were removed. The presence or absence of each uATG of *D. melanogaster* was determined at orthologous sites in *D. simulans* based on multiple genome alignments, and vice versa. The conserved uORF was defined as a uORF where its uATG was present in both *D. melanogaster* and the corresponding orthologous positions of *D. simulans*. For each protein-coding gene, we only considered the canonical transcript. For transcripts containing multiple uORFs, we defined the uORF that showed the highest TE as the dominant uORF. The branch length score (BLS) of the uATGs was calculated as we previously described¹⁸.

Fly materials and general raising conditions

The *sim4* strain of *D. simulans* was used to generate all the libraries of *D. simulans* in this study. All flies were raised on standard corn medium and grown in 12 h light: 12 h dark cycles at 25 °C for general conditions or at 29 °C for specific experimental design. The samples of embryos at different stages, larva, pupa, bodies and heads were collected following a previous protocol²³.

Processing *Drosophila* mRNA-Seq and Ribo-Seq data

Ribo-Seq and matched mRNA-Seq libraries for different developmental stages and tissues of *D. simulans* were constructed as we previously described²³ and sequenced on an Illumina HiSeq-2500 sequencer (run type: single-end; read length: 50 nt) according to the manufacturer's protocol. The 3' adaptor sequences (TGGAATTCTCGGGTGCCAAGG) were trimmed using Cutadapt 3.0 with default parameters¹¹⁰, and the NGS reads were mapped to the genomes of yeast, *Wolbachia*, *Drosophila* viruses and the sequences of tRNAs, ribosomal RNAs (rRNAs), small nuclear RNAs

(snRNAs) or small nucleolar RNAs (snoRNAs) of *D. melanogaster* (FlyBase r6.04) and *D. simulans* (FlyBase r1.3) using Bowtie2 version 2.2.3¹¹¹ with the parameters -p8 -local -k1. The mapped reads of these genomes/sequences were further removed in the downstream analysis.

After filtering, the mRNA-Seq and Ribo-Seq reads were mapped to the reference genomes of *D. melanogaster* (FlyBase, r6.04) and *D. simulans* (NCBI, ASM438218v1) respectively using the Spliced Transcripts Alignment to a Reference (STAR) algorithm¹¹². For Ribo-Seq reads, we assigned a mapped RPF (27-34 nt in length) to its P-site using the psite script from Plastid¹¹³. The uniquely mapped reads were extracted and then mapped to the CDS of *D. melanogaster* and *D. simulans* respectively using STAR¹¹². The P-sites of RPF or mRNA-Seq reads that overlapped with a CDS were counted separately. The reads that were not mapped to CDSs were then mapped to the 5'UTRs of *D. melanogaster* and *D. simulans*. The P-sites of RPF or mRNA-Seq reads that overlapped with uORFs were counted separately.

The TE of a given uORF or CDS was calculated as the $\text{RPKM}_{\text{P-site}}/\text{RPKM}_{\text{mRNA}}$ ratio. For a few uORFs with $\text{RPKM}_{\text{mRNA}} = 0$ but $\text{RPKM}_{\text{P-site}} > 0$, a pseudocount of 0.1 was added to both $\text{RPKM}_{\text{mRNA}}$ and $\text{RPKM}_{\text{P-site}}$ to avoid dividing a positive value by zero. In each sample, expressed genes were defined as the genes with a CDS $\text{RPKM}_{\text{mRNA}} > 0.1$ in both species, and translated uORFs were defined as uORFs with a TE > 0.1.

Testing the statistical significance of the difference in the interspecific TE change between a uORF and its downstream CDS

For this analysis, we adopted the methods developed by Zhang *et al.*²³. Briefly, for the samples of female bodies or male bodies of *D. melanogaster* with biological replicates, we obtained the $\log_2(\text{TE})$ and SE of the $\log_2(\text{TE})$ of a uORF or CDS by contrasting the RPF counts against mRNA-Seq read counts using DESeq2. Then, we fitted the SE values against the normalized mRNA counts and $\log_2(\text{TE})$ values using the gam function in the R package mgcv, with a log link to obtain the SE ~ mRNA counts + $\log_2(\text{TE})$ functions. For other samples without biological replicates, we estimated the SE of the $\log_2(\text{TE})$ for a feature (CDS or uORF) by applying the fitted functions obtained based on the biological replicates of female and male bodies to the observed mRNA counts and $\log_2(\text{TE})$ values. We identified uORFs whose TEs differed significantly between the paired samples of *D. melanogaster* and *D. simulans* by testing whether the value obtained for $\log_2(\beta_u) = \log_2(\text{TE}_{\text{uORF},\text{sim}}) - \log_2(\text{TE}_{\text{uORF},\text{mel}})$ was significantly different from 0. Based on the SE of the $\log_2(\text{TE})$ derived as described above, the SE of the $\log_2(\beta_u)$ can be derived as follows:

$$\text{SE}_{\log_2(\beta_u)} = \sqrt{(\text{SE}_{\log_2 \text{TE}_{\text{uORF},\text{sim}}})^2 + (\text{SE}_{\log_2 \text{TE}_{\text{uORF},\text{mel}}})^2}$$

As the Wald statistic, $\frac{\log_2(\beta_u)}{\text{SE}_{\log_2(\beta_u)}}$, follows a standard normal distribution under the null hypothesis of $\log_2(\beta_u) = 0$, we calculated the *P* value as follows: $2 \cdot (1 - \Phi(|\frac{\log_2(\beta_u)}{\text{SE}_{\log_2(\beta_u)}}|))$. The identification of CDSs whose TEs differed significantly between the paired samples of *D. melanogaster* and *D. simulans* was similar to the procedures above.

Then, we defined $\gamma = \frac{\beta_u}{\beta_c}$ (where $\beta_c = \text{TE}_{\text{CDS},\text{sim}} / \text{TE}_{\text{CDS},\text{mel}}$) and tested whether $\log_2(\gamma)$ was significantly different from 0 to determine whether the magnitude of interspecific TE changes in CDSs and uORFs was significantly different. We obtained $\log_2 \text{TE}_{\text{uORF},\text{sim}}$, $\log_2 \text{TE}_{\text{uORF},\text{mel}}$, $\log_2 \text{TE}_{\text{CDS},\text{sim}}$, and $\log_2 \text{TE}_{\text{CDS},\text{mel}}$ as described above and estimated $\text{SE}_{\log_2 \text{TE}_{\text{uORF},\text{sim}}}$, $\text{SE}_{\log_2 \text{TE}_{\text{uORF},\text{mel}}}$, $\text{SE}_{\log_2 \text{TE}_{\text{CDS},\text{sim}}}$, and $\text{SE}_{\log_2 \text{TE}_{\text{CDS},\text{mel}}}$ based on the biological replicates of female and male bodies of *D. melanogaster*. Finally, $\log_2(\gamma)$ also follows a normal distribution with SE denoted as $\text{SE}_{\log_2(\gamma)} = \sqrt{(\text{SE}_{\log_2 \text{TE}_{\text{uORF},\text{sim}}})^2 + (\text{SE}_{\log_2 \text{TE}_{\text{uORF},\text{mel}}})^2 + (\text{SE}_{\log_2 \text{TE}_{\text{CDS},\text{sim}}})^2 + (\text{SE}_{\log_2 \text{TE}_{\text{CDS},\text{mel}}})^2}$

As the Wald statistic, $\frac{\log_2(\gamma)}{\text{SE}_{\log_2(\gamma)}}$, follows a standard normal distribution under the null hypothesis that $\log_2(\beta_u) = 0$, we calculated the *P* value as follows: $2 \cdot (1 - \Phi(|\frac{\log_2(\gamma)}{\text{SE}_{\log_2(\gamma)}}|))$.

Knocking out a uORF with CRISPR-Cas9 technology in *D. melanogaster*

We searched for possible sgRNA target sites near the uATG start codon of a uORF using the Benchling website (<https://www.benchling.com/crispr/>) to design optimal single guide RNA (sgRNA) sequences with high specificity and low off-target effects. We then synthesized single-stranded complementary DNAs (ssDNAs) and annealed them to obtain double-stranded DNA (dsDNA), which served as the template for sgRNA expression. The template sequences of the sgRNA used for *bcd* uATG-KO are listed in Table S6. The dsDNA was then ligated into the BbsI-digested pU6B vector. The pU6B-sgRNA plasmid was purified and injected into the embryos of transgenic Cas9 flies collected within one hour of laying at the Tsinghua Fly Center as described in Ni *et al.* ¹¹⁴. The injected embryos were kept at 25°C and 60% humidity until adulthood (G0). The G0 adult flies that hatched from injected embryos were individually crossed with other flies (*y sc v*) to increase the number of offspring. Then, the F1 progeny were crossed with flies carrying an appropriate balancer (*Dr, e/TM3, Sb*). After F2 spawning, the F1 individuals were screened for mutations of interest by genotyping. The primers used for genotyping are listed in Table S6. The F2 progeny whose parents showed positive genotyping results were then screened for the *yellow*[−] gene to separate the chromosome carrying *nos-Cas9*. The screened F2 males were crossed with the flies containing the same balancer as above. After F2 genotyping, the progeny (F3) of positive F2 individuals showing the same mutation status were crossed individually to generate homozygous mutants in the F4 generation. The original homozygous mutants were sequentially outcrossed with *w*¹¹¹⁸ flies for 9 generations to purify the genetic background (Fig. S22B).

Ribosome fraction analysis by sucrose gradient fractionation and RT-qPCR

The 0–2 h embryos obtained from *w*¹¹¹⁸, *uKO1/uKO1*, and *uKO2/uKO2* mutants raised at 25°C and 29°C were collected and homogenized in a Dounce homogenizer with lysis buffer [50 mM Tris pH 7.5, 150 mM NaCl, 5 mM MgCl₂, 1% Triton X-100, 2 mM dithiothreitol (DTT), 20 U/ml Superscript II (Ambion), 0.5 tablets of proteinase inhibitor (Roche), 100 µg/ml emetine (Sigma Aldrich), and 50 µM guanosine 5'-[β,γ-imido]triphosphate trisodium salt hydrate (GMP-PNP) (Sigma Aldrich)] at 4°C. The lysates were clarified by centrifugation at 4°C and 20,000×g for 8 min, and the supernatants were transferred to new 1.5 ml tubes. 10–45% sucrose gradients were prepared in buffer (250 mM NaCl, 50 mM Tris pH 7.5, 15 mM MgCl₂, 0.5 mM DTT, 12 U/ml RNaseOUT, 0.5 tablets of protease inhibitor, and 20 µg/ml emetine) using a Gradient Master (Biocomp Instruments) in ULTRA-CLEAR Thinwall Tubes (Beckman Coulter). A sample volume of up to 500 µl was applied to the top of each gradient. After ultracentrifugation with a Hitachi P40ST rotor at 35,000 × rpm for 3 h at 4°C, the monosome and polysome fractions were collected, flash-frozen in liquid nitrogen, and stored at −80°C until further use.

The RNA in the monosome and polysome fractions was extracted separately using TRIzol reagent (Life Technologies, Inc.) and chloroform (Beijing Chemical Works) following the manufacturer's instructions and were reverse transcribed into cDNA using the PrimeScript™ II 1st Strand cDNA Synthesis Kit (Takara). RT-qPCR analysis of *bcd* cDNA and its targets was performed using PowerUp™ SYBR™ Green Master Mix (Thermo Fisher) following the manufacturer's instructions with *rp49* as an internal control. The primer sequences employed for RT-qPCR are listed in Table S6. For each sample, the ratio of *bcd* mRNA abundance in the polysome fraction to that in the monosome fraction was calculated as the P-to-M ratio. Six biological replicates were performed for each sample.

Dual-luciferase reporter assays

The wild-type (WT) 5'UTR of *bcd* was cloned from cDNA by PCR, and uATG mutations were introduced into 5'UTR using specific amplification primers (listed in Table S6). The WT and mutated 5'UTR sequences were ligated into a linearized reporter plasmid (psiCHECK-2 vector, Promega). The whole sequence of all the plasmids was validated by Sanger sequencing.

Drosophila S2 cells were cultured in Schneider's Insect Medium (Sigma) plus 10% (by volume) heat-inactivated fetal bovine serum, 100 U/ml penicillin and 100 µg/ml streptomycin (Thermo Fisher) at 25°C without CO₂ for 24 h to reach 1–2×10⁶ cells/ml before further treatments. Plasmid transfection was conducted with Lipofectamine 3000 (L3000001, Thermo Fisher) according to the supplier's protocol. The *Renilla* luciferase activity associated with WT or uORF-mutated 5'UTRs was measured according to the manual of the Dual-Luciferase Reporter Assay System (Promega) 32 h after transfection and was normalized to the activity of firefly luciferase.

mRNA-Seq in the embryos of mutant and WT flies

We collected 0–2 h, 2–6 h, 6–12 h, and 12–24 h embryos of *w*¹¹¹⁸ and uKO2/uKO2 mutants raised at 25°C and 29°C and conducted mRNA-Seq. Library construction and sequencing with PE150 were conducted by Annoroad on the Illumina Nova6000 platform. Two biological replicates were sequenced for each sample. The clean data were mapped to the reference genome of *D. melanogaster* (FlyBase, r6.04) using STAR ¹¹²[\[112\]](#). Reads mapped to the exons of each gene were tabulated with htseq-count ¹¹⁵[\[115\]](#). The differentially expressed genes were identified using DESeq2 ¹¹⁶[\[116\]](#). The Gene Ontology (GO) analyses were conducted using the “clusterProfiler” package in R ¹¹⁷[\[117\]](#).

Measurement of embryo hatchability

We collected embryos from WT and mutant flies and manually seeded them in vials containing standard corn medium at a density of 30 embryos per vial and 20 vials per strain. The embryos were cultivated under standard conditions (60% humidity, 12 h light: 12 h dark cycles at 25°C). We counted the number of pupae in each vial after the completion of pupation and calculated the corresponding hatching rate = $\text{number}_{\text{pupa}} / 30$

Quantification of offspring number per female fly

Newly hatched virgins were picked out and allowed to mature for two days in separate vials. They were then mated by placing one virgin female with three male flies for two days. After that, each female parent was transferred to a new vial to count the offspring number produced in each 10-day period at 25°C and 29°C. All assays were performed with 20 females per genotype.

Measurement of starvation resistance in adult flies

We selected 3- to 5-day-old adult males and females and placed them in the starvation medium (1.5% agar), with 10 flies per vial and 10 vials for both males and females from each strain. We made observations every 6 h or 12 h to count the number of deaths under starvation conditions until all flies had starved to death. The survival curves were plotted by the ggsurvplot package in R.

mRNA-Seq and Ribo-Seq data analysis in primates

The mRNA-Seq and Ribo-Seq data from the brains, livers, and testes of humans and macaques were downloaded from reference ⁷³[\[73\]](#) with accession number E-MTAB-7247 (ArrayExpress). The mRNA-Seq and Ribo-Seq data of human lymphoblastoid cell lines (LCL) from Yoruba individuals were downloaded from references ⁷⁵[\[75\]](#), ⁷⁶[\[76\]](#) under accession numbers GSE61742 and E-GEUV-1. The matched mRNA-Seq and Ribo-Seq libraries of 69 individuals were used.

The uORF annotation and downstream analysis procedures for the human and macaque data were similar to those applied in *Drosophila* as described above. The differential analysis of translational efficiency in humans and macaques was conducted by Xtail ¹¹⁸[\[23\]](#). In each pair of human-macaque samples, expressed genes were defined as the genes with a CDS RPKM_{mRNA} > 0.1 in both species. The translated uORFs in a sample were defined as uORFs with a TE > 0.1. For the human cell line data, expressed genes were defined as genes with a mean CDS RPKM_{mRNA} > 0.1 across the cell lines, and translated uORFs were defined as uORFs with a mean TE > 0.1

Acknowledgements

We thank Drs. Wei Xie, Weiwei Zhai, and Xionglei He for constructive comments and suggestions. This work was supported by grants from the Ministry of Science and Technology of the People's Republic of China (2022YFE0132000), the Yunnan Provincial Science and Technology Project at Southwest United Graduate School (202302A0370006), the National Natural Science Foundation of China (32070597) and the Natural Science Foundation of Beijing (5212006). We thank the National Center for Protein Sciences at Peking University for their technical assistance. Some of the analyses were performed on the High-Performance Computing Platform of the Center for Life Sciences.

Declaration of interests

The authors declare no competing interests.

Data availability statement

All deep-sequencing data generated in this study, including single-ended mRNA-Seq and Ribo-Seq data of 10 developmental stages and tissues of *Drosophila simulans* and paired-end mRNA-Seq data of 0-2 h, 2-6 h, 6-12 h, and 12-24 h *Drosophila melanogaster* embryos, were deposited in the China National Genomics Data Center Genome Sequence Archive (GSA) under accession numbers CRA003198, CRA007425, and CRA007426. The mRNA-Seq and Ribo-Seq data for the different developmental stages and tissues of *Drosophila melanogaster* were published in our previous paper ²³[\[23\]](#) and were deposited in the Sequence Read Archive (SRA) under accession number SRP067542.

References

1. Buccitelli C., Selbach M. (2020) **mRNAs, proteins and the emerging principles of gene expression control** *Nature Reviews Genetics* **21**:630–644 <https://doi.org/10.1038/s41576-020-0258-4>
2. Lee T.I., Young R.A. (2013) **Transcriptional regulation and its misregulation in disease** *Cell* **152**:1237–1251 <https://doi.org/10.1016/j.cell.2013.02.014>
3. MacNeil L.T., Walhout A.J.M. (2011) **Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression** *Genome Research* **21**:645–657 <https://doi.org/10.1101/gr.097378.109>
4. Hill M.S., Vande Zande P., Wittkopp P.J. (2021) **Molecular and evolutionary processes generating variation in gene expression** *Nat Rev Genet* **22**:203–215 <https://doi.org/10.1038/s41576-020-00304-w>
5. Vogel C. (2013) **Protein Expression Under Pressure** *Science* **342**:1052–1053 <https://doi.org/10.1126/science.1247833>
6. Signor S.A., Nuzhdin S.V. (2018) **The Evolution of Gene Expression in cis and trans** *Trends Genet* **34**:532–544 <https://doi.org/10.1016/j.tig.2018.03.007>
7. Artieri C.G., Fraser H.B. (2014) **Evolution at two levels of gene expression in yeast** *Genome Res* **24**:411–421 <https://doi.org/10.1101/gr.165522.113>
8. McManus C.J., May G.E., Spealman P., Shteyman A. (2014) **Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast** *Genome Res* **24**:422–430 <https://doi.org/10.1101/gr.164996.113>
9. Wang Z., Sun X., Zhao Y., Guo X., Jiang H., Li H., Gu Z. (2015) **Evolution of gene regulation during transcription and translation** *Genome Biol Evol* **7**:1155–1167 <https://doi.org/10.1093/gbe/evv059>
10. Khan Z., Ford M.J., Cusanovich D.A., Mitrano A., Pritchard J.K., Gilad Y. (2013) **Primate transcript and protein expression levels evolve under compensatory selection pressures** *Science* **342**:1100–1104 <https://doi.org/10.1126/science.1242379>
11. Wang S.H., Hsiao C.J., Khan Z., Pritchard J.K. (2018) **Post-translational buffering leads to convergent protein expression levels between primates** *Genome Biol* **19** <https://doi.org/10.1186/s13059-018-1451-z>
12. Schrimpf S.P. *et al.* (2009) **Comparative functional analysis of the Caenorhabditis elegans and Drosophila melanogaster proteomes** *PLoS Biol* **7** <https://doi.org/10.1371/journal.pbio.1000048>
13. Kusnadi E.P., Timpone C., Topisirovic I., Larsson O., Furic L. (2022) **Regulation of gene expression via translational buffering** *Biochim Biophys Acta Mol Cell Res* **1869** <https://doi.org/10.1016/j.bbamcr.2021.119140>

14. Laurent J.M. *et al.* (2010) **Protein abundances are more conserved than mRNA abundances across diverse taxa** *Proteomics* **10**:4209–4212 <https://doi.org/10.1002/pmic.201000327>
15. Teixeira F.K., Lehmann R. (2019) **Translational Control during Developmental Transitions** *Cold Spring Harb Perspect Biol* **11** <https://doi.org/10.1101/cshperspect.a032987>
16. Jackson R.J., Hellen C.U., Pestova T.V. (2010) **The mechanism of eukaryotic translation initiation and principles of its regulation** *Nat Rev Mol Cell Biol* **11**:113–127 <https://doi.org/10.1038/nrm2838>
17. Sonenberg N., Hinnebusch A.G. (2009) **Regulation of translation initiation in eukaryotes: mechanisms and biological targets** *Cell* **136**:731–745 <https://doi.org/10.1016/j.cell.2009.01.042>
18. Zhang H., Wang Y., Wu X., Tang X., Wu C., Lu J. (2021) **Determinants of genome-wide distribution and evolution of uORFs in eukaryotes** *Nature Communications* **12** <https://doi.org/10.1038/s41467-021-21394-y>
19. Zhang H., Wang Y., Lu J. (2019) **Function and Evolution of Upstream ORFs in Eukaryotes** *Trends Biochem Sci* **44**:782–794 <https://doi.org/10.1016/j.tibs.2019.03.002>
20. Neafsey D.E., Galagan J.E. (2007) **Dual modes of natural selection on upstream open reading frames** *Mol Biol Evol* **24**:1744–1751 <https://doi.org/10.1093/molbev/msm093>
21. Churbanov A., Rogozin I.B., Babenko V.N., Ali H., Koonin E.V. (2005) **Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes** *Nucleic Acids Res* **33**:5512–5520 <https://doi.org/10.1093/nar/gki847>
22. Resch A.M., Ogurtsov A.Y., Rogozin I.B., Shabalina S.A., Koonin E.V. (2009) **Evolution of alternative and constitutive regions of mammalian 5'UTRs** *BMC Genomics* **10** <https://doi.org/10.1186/1471-2164-10-162>
23. Zhang H., Dou S., He F., Luo J., Wei L., Lu J. (2018) **Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during *Drosophila* development** *PLoS biology* **16**
24. Malzer E., Szajewska-Skuta M., Dalton L.E., Thomas S.E., Hu N., Skaer H., Lomas D.A., Crowther D.C., Marciniak S.J. (2013) **Coordinate regulation of eIF2alpha phosphorylation by PPP1R15 and GCN2 is required during *Drosophila* development** *J Cell Sci* **126**:1406–1415 <https://doi.org/10.1242/jcs.117614>
25. Komonyi O., Papai G., Enunlu I., Muratoglu S., Pankotai T., Kopitova D., Maroy P., Udvardy A., Boros I. (2005) **DTL, the *Drosophila* homolog of PIMT/Tgs1 nuclear receptor coactivator-interacting protein/RNA methyltransferase, has an essential role in development** *J Biol Chem* **280**:12397–12404 <https://doi.org/10.1074/jbc.M409251200>
26. Medenbach J., Seiler M., Hentze M.W. (2011) **Translational control via protein-regulated upstream open reading frames** *Cell* **145**:902–913 <https://doi.org/10.1016/j.cell.2011.05.005>
27. Chen J., Tresenrider A., Chia M., McSwiggen D.T., Spedale G., Jorgensen V., Liao H., van Werven F.J., Unal E. (2017) **Kinetochores inactivation by expression of a repressive mRNA** *Elife* **6** <https://doi.org/10.7554/eLife.27417>

28. Cheng Z., Otto G.M., Powers E.N., Keskin A., Mertins P., Carr S.A., Jovanovic M., Brar G.A. (2018) **Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis** *Cell* **172**:910–923 <https://doi.org/10.1016/j.cell.2018.01.035>
29. Kurihara Y., Makita Y., Kawashima M., Fujita T., Iwasaki S., Matsui M. (2018) **Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in Arabidopsis** *Proc Natl Acad Sci U S A* **115**:7831–7836 <https://doi.org/10.1073/pnas.1804971115>
30. Yang Y.F., Zhang X., Ma X., Zhao T., Sun Q., Huan Q., Wu S., Du Z., Qian W. (2017) **Trans-splicing enhances translational efficiency in *C. elegans*** *Genome Res* **27**:1525–1535 <https://doi.org/10.1101/gr.202150.115>
31. Wiestner A., Schlemper R.J., van der Maas A.P., Skoda R.C. (1998) **An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia** *Nat Genet* **18**:49–52 <https://doi.org/10.1038/ng0198-49>
32. Liu L., Dilworth D., Gao L., Monzon J., Summers A., Lassam N., Hogg D. (1999) **Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma** *Nat Genet* **21**:128–132 <https://doi.org/10.1038/5082>
33. Wen Y. *et al.* (2009) **Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis** *Nat Genet* **41**:228–233 <https://doi.org/10.1038/ng.276>
34. Calvo S.E., Pagliarini D.J., Mootha V.K. (2009) **Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans** *Proc Natl Acad Sci U S A* **106**:7507–7512 <https://doi.org/10.1073/pnas.0810916106>
35. Lee D.S.M., Park J., Kromer A., Baras A., Rader D.J., Ritchie M.D., Ghanem L.R., Barash Y. (2021) **Disrupting upstream translation in mRNAs is associated with human disease** *Nat Commun* **12** <https://doi.org/10.1038/s41467-021-21812-1>
36. Costa-Mattioli M., Walter P. (2020) **The integrated stress response: From mechanism to disease** *Science* **368** <https://doi.org/10.1126/science.aat5314>
37. Hinnebusch A.G. (2005) **Translational regulation of GCN4 and the general amino acid control of yeast** *Annu Rev Microbiol* **59**:407–450 <https://doi.org/10.1146/annurev.micro.59.031805.133833>
38. Lu P.D., Harding H.P., Ron D. (2004) **Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response** *J Cell Biol* **167**:27–33 <https://doi.org/10.1083/jcb.200408003>
39. Bohlen J., Harbrecht L., Blanco S., Clemm von Hohenberg K., Fenzl K., Kramer G., Bukau B., Teleman A.A. (2020) **DENR promotes translation reinitiation via ribosome recycling to drive expression of oncogenes including ATF4** *Nat Commun* **11** <https://doi.org/10.1038/s41467-020-18452-2>
40. Vasudevan D., Neuman S.D., Yang A., Lough L., Brown B., Bashirullah A., Cardozo T., Ryoo H.D. (2020) **Translational induction of ATF4 during integrated stress response requires noncanonical initiation factors eIF2D and DENR** *Nat Commun* **11** <https://doi.org/10.1038/s41467-020-18453-1>

41. Vattam K.M., Wek R.C. (2004) **Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells** *Proc Natl Acad Sci U S A* **101**:11269–11274 <https://doi.org/10.1073/pnas.0400541101>
42. Andreev D.E., O'Connor P.B., Fahey C., Kenny E.M., Terenin I.M., Dmitriev S.E., Cormican P., Morris D.W., Shatsky I.N., Baranov P.V. (2015) **Translation of 5' leaders is pervasive in genes resistant to eIF2 repression** *Elife* **4** <https://doi.org/10.7554/eLife.03971>
43. Andreev D.E., Arnold M., Kiniry S.J., Loughran G., Michel A.M., Rachinskii D., Baranov P.V. (2018) **TASEP modelling provides a parsimonious explanation for the ability of a single uORF to derepress translation during the integrated stress response** *Elife* **7** <https://doi.org/10.7554/eLife.32563>
44. Kozak M. (1989) **The scanning model for translation: an update** *J Cell Biol* **108**:229–241 <https://doi.org/10.1083/jcb.108.2.229>
45. Johnstone T.G., Bazzini A.A., Giraldez A.J. (2016) **Upstream ORFs are prevalent translational repressors in vertebrates** *EMBO J* **35**:706–723 <https://doi.org/10.15252/embj.201592759>
46. Hinnebusch A.G., Ivanov I.P., Sonenberg N. (2016) **Translational control by 5'-untranslated regions of eukaryotic mRNAs** *Science* **352**:1413–1416 <https://doi.org/10.1126/science.aad9868>
47. Morris D.R., Geballe A.P. (2000) **Upstream open reading frames as regulators of mRNA translation** *Mol Cell Biol* **20**:8635–8642 <https://doi.org/10.1128/MCB.20.23.8635-8642.2000>
48. Young S.K., Wek R.C. (2016) **Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response** *J Biol Chem* **291**:16927–16935 <https://doi.org/10.1074/jbc.R116.733899>
49. Dever T.E., Feng L., Wek R.C., Cigan A.M., Donahue T.F., Hinnebusch A.G. (1992) **Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast** *Cell* **68**:585–596 [https://doi.org/10.1016/0092-8674\(92\)90193-g](https://doi.org/10.1016/0092-8674(92)90193-g)
50. Palam L.R., Baird T.D., Wek R.C. (2011) **Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation** *J Biol Chem* **286**:10939–10949 <https://doi.org/10.1074/jbc.M110.216093>
51. Baird T.D., Palam L.R., Fusakio M.E., Willy J.A., Davis C.M., McClintick J.N., Anthony T.G., Wek R.C. (2014) **Selective mRNA translation during eIF2 phosphorylation induces expression of IBTKalpha** *Mol Biol Cell* **25**:1686–1697 <https://doi.org/10.1091/mbc.E14-02-0704>
52. Chen Y.J., Tan B.C., Cheng Y.Y., Chen J.S., Lee S.C. (2010) **Differential regulation of CHOP translation by phosphorylated eIF4E under stress conditions** *Nucleic Acids Res* **38**:764–777 <https://doi.org/10.1093/nar/gkp1034>
53. Fraser H.B., Hirsh A.E., Giaever G., Kumm J., Eisen M.B. (2004) **Noise minimization in eukaryotic gene expression** *PLoS Biol* **2** <https://doi.org/10.1371/journal.pbio.0020137>
54. Thattai M., van Oudenaarden A. (2001) **Intrinsic noise in gene regulatory networks** *Proc Natl Acad Sci U S A* **98**:8614–8619 <https://doi.org/10.1073/pnas.151588598>

55. Wu H.W., Fajiculay E., Wu J.F., Yan C.S., Hsu C.P., Wu S.H. (2022) **Noise reduction by upstream open reading frames** *Nat Plants* **8**:474–480 <https://doi.org/10.1038/s41477-022-01136-8>
56. Bottorff T.A., Park H., Geballe A.P., Subramaniam A.R. (2022) **Translational buffering by ribosome stalling in upstream open reading frames** *PLoS Genet* **18** <https://doi.org/10.1371/journal.pgen.1010460>
57. Ciandrini L., Stansfield I., Romano M.C. (2010) **Role of the particle's stepping cycle in an asymmetric exclusion process: a model of mRNA translation** *Phys Rev E Stat Nonlin Soft Matter Phys* **81** <https://doi.org/10.1103/PhysRevE.81.051904>
58. Reuveni S., Meilijson I., Kupiec M., Rupp E., Tuller T. (2011) **Genome-scale analysis of translation elongation with a ribosome flow model** *PLoS Comput Biol* **7** <https://doi.org/10.1371/journal.pcbi.1002127>
59. von der Haar T. (2012) **Mathematical and Computational Modelling of Ribosomal Movement and Protein Synthesis: an overview** *Comput Struct Biotechnol J* **1** <https://doi.org/10.5936/csbj.201204002>
60. Zhao Y.B., Krishnan J. (2014) **mRNA translation and protein synthesis: an analysis of different modelling methodologies and a new PBN based approach** *BMC Syst Biol* **8** <https://doi.org/10.1186/1752-0509-8-25>
61. Tamura K., Subramanian S., Kumar S. (2004) **Temporal Patterns of Fruit Fly (*Drosophila*) Evolution Revealed by Mutation Clocks** *Molecular Biology and Evolution* **21**:36–44 <https://doi.org/10.1093/molbev/msg236>
62. Kronja I., Yuan B., Eichhorn S.W., Dzek K., Krijgsveld J., Bartel D.P., Orr-Weaver T.L. (2014) **Widespread changes in the posttranscriptional landscape at the *Drosophila* oocyte-to-embryo transition** *Cell Rep* **7**:1495–1508 <https://doi.org/10.1016/j.celrep.2014.05.002>
63. Qin X., Ahn S., Speed T.P., Rubin G.M. (2007) **Global analyses of mRNA translational control during early *Drosophila* embryogenesis** *Genome Biol* **8** <https://doi.org/10.1186/gb-2007-8-4-r63>
64. Berleth T., Burri M., Thoma G., Bopp D., Richstein S., Frigerio G., Noll M., Nusslein-Volhard C. (1988) **The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo** *EMBO J* **7**:1749–1756
65. Driever W., Nusslein-Volhard C. (1988) **A gradient of bicoid protein in *Drosophila* embryos** *Cell* **54**:83–93 [https://doi.org/10.1016/0092-8674\(88\)90182-1](https://doi.org/10.1016/0092-8674(88)90182-1)
66. Driever W., Nusslein-Volhard C. (1988) **The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner** *Cell* **54**:95–104 [https://doi.org/10.1016/0092-8674\(88\)90183-3](https://doi.org/10.1016/0092-8674(88)90183-3)
67. Abdelmohsen K. *et al.* (2014) **7SL RNA represses p53 translation by competing with HuR** *Nucleic Acids Res* **42**:10099–10111 <https://doi.org/10.1093/nar/gku686>
68. Panda A.C. *et al.* (2016) **Novel RNA-binding activity of MYF5 enhances Ccnd1/Cyclin D1 mRNA translation during myogenesis** *Nucleic Acids Res* **44**:2393–2408 <https://doi.org/10.1093/nar/gkw023>

69. Panda A.C., Martindale J.L., Gorospe M. (2017) **Polysome Fractionation to Analyze mRNA Distribution Profiles** *Bio Protoc* **7** <https://doi.org/10.21769/BioProtoc.2126>
70. Amourda C., Chong J., Saunders T.E. (2018) **MicroRNAs buffer genetic variation at specific temperatures during embryonic development** *bioRxiv* **444810** <https://doi.org/10.1101/444810>
71. Lu G., Zhao Y., Chen Q., Lin P., Tang T., Tang Z., Liufu Z., Wu C.-I. (2021) **When development is constantly but weakly perturbed - Canalization by microRNAs** *bioRxiv* <https://doi.org/10.1101/2021.09.04.458966>
72. Li X.Y. *et al.* (2008) **Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm** *PLoS Biol* **6** <https://doi.org/10.1371/journal.pbio.0060027>
73. Wang Z.Y. *et al.* (2020) **Transcriptome and translome co-evolution in mammals** *Nature* **588**:642–647 <https://doi.org/10.1038/s41586-020-2899-z>
74. Hedges S.B., Dudley J., Kumar S. (2006) **TimeTree: a public knowledge-base of divergence times among organisms** *Bioinformatics* **22**:2971–2972 <https://doi.org/10.1093/bioinformatics/btl505>
75. Battle A., Khan Z., Wang S.H., Mitrano A., Ford M.J., Pritchard J.K., Gilad Y. (2015) **Genomic variation. Impact of regulatory variation from RNA to protein** *Science* **347**:664–667 <https://doi.org/10.1126/science.1260793>
76. Lappalainen T. *et al.* (2013) **Transcriptome and genome sequencing uncovers functional variation in humans** *Nature* **501**:506–511 <https://doi.org/10.1038/nature12531>
77. Stein K.C., Frydman J. (2019) **The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis** *Journal of Biological Chemistry* **294**:2076–2084 <https://doi.org/10.1074/jbc.REV118.002814>
78. Sherman M.Y., Qian S.B. (2013) **Less is more: improving proteostasis by translation slow down** *Trends Biochem Sci* **38**:585–591 <https://doi.org/10.1016/j.tibs.2013.09.003>
79. Waddington C.H. (1942) **CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS** *Nature* **150**:563–565 <https://doi.org/10.1038/150563a0>
80. Payne J.L., Wagner A. (2015) **Mechanisms of mutational robustness in transcriptional regulation** *Front Genet* **6** <https://doi.org/10.3389/fgene.2015.00322>
81. Denby C.M., Im J.H., Yu R.C., Pesce C.G., Brem R.B. (2012) **Negative feedback confers mutational robustness in yeast transcription factor regulation** *Proc Natl Acad Sci U S A* **109**:3874–3878 <https://doi.org/10.1073/pnas.1116360109>
82. Jarosz D.F., Taipale M., Lindquist S. (2010) **Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms** *Annu Rev Genet* **44**:189–216 <https://doi.org/10.1146/annurev.genet.40.110405.090412>
83. Ebert M.S., Sharp P.A. (2012) **Roles for microRNAs in conferring robustness to biological processes** *Cell* **149**:515–524 <https://doi.org/10.1016/j.cell.2012.04.005>

84. Lu G.A., Zhang J., Zhao Y., Chen Q., Lin P., Tang T., Tang Z., Wen H., Liufu Z., Wu C.I. (2023) **Canalization of Phenotypes-When the Transcriptome is Constantly but Weakly Perturbed** *Mol Biol Evol* **40** <https://doi.org/10.1093/molbev/msad005>
85. Alon U. (2007) **Network motifs: theory and experimental approaches** *Nat Rev Genet* **8**:450–461 <https://doi.org/10.1038/nrg2102>
86. Somogyvari M., Khatatneh S., Soti C. (2022) **Hsp90: From Cellular to Organismal Proteostasis** *Cells* **11** <https://doi.org/10.3390/cells11162479>
87. Zabinsky R.A., Mason G.A., Queitsch C., Jarosz D.F. (2019) **It's not magic - Hsp90 and its effects on genetic and epigenetic variation** *Semin Cell Dev Biol* **88**:21–35 <https://doi.org/10.1016/j.semcdb.2018.05.015>
88. Cho P.F., Poulin F., Cho-Park Y.A., Cho-Park I.B., Chicoine J.D., Lasko P., Sonenberg N. (2005) **A new paradigm for translational control: inhibition via 5'-3' mRNA tethering by Bicoid and the eIF4E cognate 4EHP** *Cell* **121**:411–423 <https://doi.org/10.1016/j.cell.2005.02.024>
89. Singh A.P., Wu P., Ryabichko S., Raimundo J., Swan M., Wieschaus E., Gregor T., Toettcher J.E. (2022) **Optogenetic control of the Bicoid morphogen reveals fast and slow modes of gap gene regulation** *Cell Rep* **38** <https://doi.org/10.1016/j.celrep.2022.110543>
90. Hannon C.E., Blythe S.A., Wieschaus E.F. (2017) **Concentration dependent chromatin states induced by the bicoid morphogen gradient** *Elife* **6** <https://doi.org/10.7554/eLife.28275>
91. Struhl G., Struhl K., Macdonald P.M. (1989) **The gradient morphogen bicoid is a concentration-dependent transcriptional activator** *Cell* **57**:1259–1273 [https://doi.org/10.1016/0092-8674\(89\)90062-7](https://doi.org/10.1016/0092-8674(89)90062-7)
92. Dubnau J., Struhl G. (1996) **RNA recognition and translational regulation by a homeodomain protein** *Nature* **379**:694–699 <https://doi.org/10.1038/379694a0>
93. Wethmar K., Begay V., Smink J.J., Zaragoza K., Wiesenthal V., Dorken B., Calkhoven C.F., Leutz A. (2010) **C/EBPbetaDeltauORF mice--a genetic model for uORF-mediated translational control in mammals** *Genes Dev* **24**:15–20 <https://doi.org/10.1101/gad.557910>
94. Miyake T. *et al.* (2023) **Minimal upstream open reading frame of Per2 mediates phase fitness of the circadian clock to day/night physiological body temperature rhythm** *Cell Rep* **42** <https://doi.org/10.1016/j.celrep.2023.112157>
95. Xing S., Chen K., Zhu H., Zhang R., Zhang H., Li B., Gao C. (2020) **Fine- tuning sugar content in strawberry** *Genome Biol* **21** <https://doi.org/10.1186/s13059-020-02146-5>
96. Zhang H., Si X., Ji X., Fan R., Liu J., Chen K., Wang D., Gao C. (2018) **Genome editing of upstream open reading frames enables translational control in plants** *Nat Biotechnol* **36**:894–898 <https://doi.org/10.1038/nbt.4202>
97. Si X., Zhang H., Wang Y., Chen K., Gao C. (2020) **Manipulating gene translation in plants by CRISPR-Cas9-mediated genome editing of upstream open reading frames** *Nat Protoc* **15**:338–363 <https://doi.org/10.1038/s41596-019-0238-3>
98. Xue C., Qiu F., Wang Y., Li B., Zhao K.T., Chen K., Gao C. (2023) **Tuning plant phenotypes by precise, graded downregulation of gene expression** *Nat Biotechnol* <https://doi.org/10.1038/s41587-023-01707-w>

99. Ivanov I.P., Shin B.S., Loughran G., Tzani I., Young-Baird S.K., Cao C., Atkins J.F., Dever T.E. (2018) **Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing** *Mol Cell* **70**:254–264 <https://doi.org/10.1016/j.molcel.2018.03.015>
100. Luo Z., Sachs M.S. (1996) **Role of an upstream open reading frame in mediating arginine-specific translational control in *Neurospora crassa*** *J Bacteriol* **178**:2172–2177 <https://doi.org/10.1128/jb.178.8.2172-2177.1996>
101. Lovett P.S., Rogers E.J. (1996) **Ribosome regulation by the nascent peptide** *Microbiol Rev* **60**:366–385 <https://doi.org/10.1128/mr.60.2.366-385.1996>
102. Vilela C., McCarthy J.E. (2003) **Regulation of fungal gene expression via short open reading frames in the mRNA 5′ untranslated region** *Mol Microbiol* **49**:859–867 <https://doi.org/10.1046/j.1365-2958.2003.03622.x>
103. Raney A., Law G.L., Mize G.J., Morris D.R. (2002) **Regulated translation termination at the upstream open reading frame in s-adenosylmethionine decarboxylase mRNA** *J Biol Chem* **277**:5988–5994 <https://doi.org/10.1074/jbc.M108375200>
104. Lin Y., May G.E., Kready H., Nazzaro L., Mao M., Spealman P., Creeger Y., McManus C.J. (2019) **Impacts of uORF codon identity and position on translation regulation** *Nucleic Acids Res* **47**:9358–9367 <https://doi.org/10.1093/nar/gkz681>
105. Meijer H.A., Thomas A.A. (2003) **Ribosomes stalling on uORF1 in the *Xenopus* Cx41 5′ UTR inhibit downstream translation initiation** *Nucleic Acids Res* **31**:3174–3184 <https://doi.org/10.1093/nar/gkg429>
106. Rosenbloom K.R. *et al.* (2015) **The UCSC Genome Browser database: 2015 update** *Nucleic Acids Res* **43**:D670–681 <https://doi.org/10.1093/nar/gku1177>
107. Chakraborty M., Chang C.H., Khost D.E., Vedanayagam J., Adrion J.R., Liao Y., Montooth K.L., Meiklejohn C.D., Larracuent A.M., Emerson J.J. (2021) **Evolution of genome structure in the *Drosophila simulans* species complex** *Genome Res* **31**:380–396 <https://doi.org/10.1101/gr.263442.120>
108. Chiaromonte F., Yap V.B., Miller W. (2002) **Scoring pairwise genomic sequence alignments** *Pac Symp Biocomput* :115–126 https://doi.org/10.1142/9789812799623_0012
109. Blanchette M. *et al.* (2004) **Aligning multiple genomic sequences with the threaded blockset aligner** *Genome Res* **14**:708–715 <https://doi.org/10.1101/gr.1933104>
110. Martin M. (2011) **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet journal* **17**:10–12
111. Langmead B., Salzberg S.L. (2012) **Fast gapped-read alignment with Bowtie 2** *Nat Methods* **9**:357–359 <https://doi.org/10.1038/nmeth.1923>
112. 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
113. Dunn J.G., Weissman J.S. (2016) **Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data** *BMC Genomics* **17** <https://doi.org/10.1186/s12864-016-3278-x>

- 114. Ni J.Q. *et al.* (2011) **A genome-scale shRNA resource for transgenic RNAi in *Drosophila*** *Nat Methods* **8**:405–407 <https://doi.org/10.1038/nmeth.1592>
- 115. Anders S., Pyl P.T., Huber W. (2015) **HTSeq--a Python framework to work with high-throughput sequencing data** *Bioinformatics* **31**:166–169 <https://doi.org/10.1093/bioinformatics/btu638>
- 116. McCarthy D.J., Chen Y., Smyth G.K. (2012) **Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation** *Nucleic Acids Res* **40**:4288–4297 <https://doi.org/10.1093/nar/gks042>
- 117. Yu G., Wang L.G., Han Y., He Q.Y. (2012) **clusterProfiler: an R package for comparing biological themes among gene clusters** *Omic* **16**:284–287 <https://doi.org/10.1089/omi.2011.0118>
- 118. Xiao Z., Zou Q., Liu Y., Yang X. (2016) **Genome-wide assessment of differential translations with ribosome profiling data** *Nat Commun* **7** <https://doi.org/10.1038/ncomms11194>

Author information

Yuanqiang Sun[#]

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China
ORCID iD: [0009-0007-6967-2234](https://orcid.org/0009-0007-6967-2234)

[#]Equal contribution

Yuange Duan[#]

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China
ORCID iD: [0000-0003-2311-9859](https://orcid.org/0000-0003-2311-9859)

[#]Equal contribution

Peixiang Gao

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China

Chenlu Liu

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China
ORCID iD: [0000-0002-8993-0145](https://orcid.org/0000-0002-8993-0145)

Kaichun Jin

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China

Shengqian Dou

Eye Institute of Shandong First Medical University, State Key Laboratory Cultivation Base, Shandong Provincial Key Laboratory of Ophthalmology, Qingdao, China

Wenxiong Tang

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China

Hong Zhang

College of Ecology, Lanzhou University, Lanzhou, China

Jian Lu

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China

ORCID iD: [0000-0002-4409-1667](https://orcid.org/0000-0002-4409-1667)

For correspondence: luj@pku.edu.cn

Editors

Reviewing Editor

Wenfeng Qian

Chinese Academy of Sciences, Beijing, China

Senior Editor

Alan Moses

University of Toronto, Toronto, Canada

Reviewer #1 (Public review):**Summary:**

The authors set out to explore the role of upstream open reading frames (uORFs) in stabilizing protein levels during *Drosophila* development and evolution. By utilizing a modified ICIER model for ribosome translation simulations and conducting experimental validations in *Drosophila* species, the study investigates how uORFs buffer translational variability of downstream coding sequences. The findings reveal that uORFs significantly reduce translational variability, which contributes to gene expression stability across different biological contexts and evolutionary timeframes.

Strengths:

- (1) The study introduces a sophisticated adaptation of the ICIER model, enabling detailed simulation of ribosomal traffic and its implications for translation efficiency.
- (2) The integration of computational predictions with empirical data through knockout experiments and transcriptome analysis in *Drosophila* provides a compelling validation of the model's predictions.
- (3) By demonstrating the evolutionary conservation of uORFs' buffering effects, the study provides insights that are likely applicable to a wide range of eukaryotes.

Weaknesses:

- (1) Although the study is technically sound, it does not clearly articulate the mechanisms through which uORFs buffer translational variability. A clearer hypothesis detailing the potential molecular interactions or regulatory pathways by which uORFs influence translational stability would enhance the comprehension and impact of the findings.

(2) The study could be further improved by a discussion regarding the evolutionary selection of uORFs. Specifically, it would be beneficial to explore whether uORFs are favored evolutionarily primarily for their role in reducing translation efficiency or for their capability to stabilize translation variability. Such a discussion would provide deeper insights into the evolutionary dynamics and functional significance of uORFs in genetic regulation.

<https://doi.org/10.7554/eLife.104074.1.sa1>

Reviewer #2 (Public review):

uORFs, short open reading frames located in the 5' UTR, are pervasive in genomes. However, their roles in maintaining protein abundance are not clear. In this study, the authors propose that uORFs act as "molecular dam", limiting the fluctuation of the translation of downstream coding sequences. First, they performed in silico simulations using an improved ICIER model, and demonstrated that uORF translation reduces CDS translational variability, with buffering capacity increasing in proportion to uORF efficiency, length, and number. Next, they analyzed the transcriptome between two related *Drosophila* species, revealing that genes with uORFs exhibit smaller fluctuations in translation between the two species and across different developmental stages within the same species. Moreover, they identified that *bicoid*, a critical gene for *Drosophila* development, contains a uORF with substantial changes in translation efficiency. Deleting this uORF in *Drosophila melanogaster* significantly affected its gene expression, hatching rates, and survival under stress condition. Lastly, by leveraging public Ribo-seq data, the authors showed that the buffering effect of uORFs is also evident between primates and within human populations. Collectively, the study advances our understanding of how uORFs regulate the translation of downstream coding sequences at the genome-wide scale, as well as during development and evolution.

The conclusions of this paper are mostly well supported by data, but some definitions and data analysis need to be clarified and extended.

(1) There are two definitions of translation efficiency (TE) in the manuscript: one refers to the number of 80S ribosomes that complete translation at the stop codon of a CDS within a given time interval, while the other is calculated based on Ribo-seq and mRNA-seq data (as described on Page 7, line 209). To avoid potential misunderstandings, please use distinct terms to differentiate these two definitions.

(2) Page 7, line 209: "The translational efficiencies (TEs) of the conserved uORFs were highly correlated between the two species across all developmental stages and tissues examined, with Spearman correlation coefficients ranging from 0.478 to 0.573 (Fig. 2A)." However, the authors did not analyze the correlation of translation efficiency of conserved CDSs between the two species, and compare this correlation to the correlation between the TEs of CDSs. These analyses will further support the authors' conclusion regarding the role of conserved uORFs in translation regulation.

(3) Page 8, line 217: "Among genes with multiple uORFs, one uORF generally emerged as dominant, displaying a higher TE than the others within the same gene (Fig. 2C)." The basis for determining dominance among uORFs is not explained and this lack of clarification undermines the interpretation of these findings.

(4) According to the simulation, the translation of uORFs should exhibit greater variability than that of CDSs. However, the authors observed significantly fewer uORFs with significant TE changes compared to CDSs. This discrepancy may be due to lower sequencing depth resulting in fewer reads mapped to uORFs. Therefore, the authors may compare this variability specifically among highly expressed genes.

(5) If possible, the author may need to use antibodies against bicoid to test the effect of ATG deletion on bicoid expression, particularly under different developmental stages or growth conditions. According to the authors' conclusions, the deletion mutant should exhibit greater variability in bicoid protein abundance. This experiment could provide strong support for the proposed mechanisms.

<https://doi.org/10.7554/eLife.104074.1.sa0>