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

Yuanqiang Sun[1,#], Yuange Duan[1,#], Peixiang Gao[1], Chenlu Liu[1], Kaichun Jin[1], Shengqian Dou[2], Wenxiong Tang[1], Hong Zhang[3], & Jian Lu[1,*]

1. State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing 100871, China
2. Eye Institute of Shandong First Medical University, State Key Laboratory Cultivation Base, Shandong Provincial Key Laboratory of Ophthalmology, Qingdao, Shandong, China
3. College of Ecology, Lanzhou University, Lanzhou, Gansu, China

# Equal contribution

* Correspondence:

Jian Lu

Email: luj@pku.edu.cn

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 translatomes 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.

Key words: Upstream open reading frames (uORFs), TASEP modeling, stabilizing selection, translational buffer, interspecific gene expression, canalization, *bicoid*

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 translatomes 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 model [43], 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 CDS [43].

We extended the original ICIER model [43] 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 regulation [56]. 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 quantitively 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 genes [18].

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 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 (ν_s = 0.3), the probability of an 80S ribosome moving to the next position within the uORF (ν_{Eu} = 0.3) and within the CDS (ν_{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 model [43], 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 (L_{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_{inm}) of the 1000 R_{in} inputs as the baseline, calculating the corresponding N_{ECm} and N_{EUm}values. We then calculated the difference in changes in N_{EC} (ΔN_{EC}) and N_{EU} (ΔN_{EU}) relative to N_{ECm} and N_{EUm} 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 $\log2(\Delta N_{EC}/\Delta N_{EU})$] 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_{uORF1} (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_{uORF2}

(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 translatome 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 translatomes of two closely related *Drosophila* species, *D. melanogaster* and *D. simulans*, which diverged approximately 5.4 million years ago [61]. We generated high-throughput sequencing data for *D. simulans*, including transcriptome (mRNA-Seq) and translatome (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* data [23], with identical embryonic stages and tissue types. This comprehensive comparative translatome 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).

216 In *D. melanogaster*, 7,259 (52.2%) genes had no uORFs, 2,687 (19.3%) had a single uORF, and 3,961
217 (28.5%) contained multiple uORFs. Among genes with multiple uORFs, one uORF generally emerged
218 as dominant, displaying a higher TE than the others within the same gene (Fig. 2C). The median TE of
219 the dominant uORF was 4.84 times that of the second-highest uORF within the same gene in *D.*
220 *melanogaster*, and the corresponding ratio was 5.21 times in *D. simulans* (Fig. 2C). To assess the
221 consistency of this dominance across different tissues and developmental stages, we identified 3,072
222 multiple-uORF genes in *D. melanogaster* with at least one translated uORF (TE > 0.1 in at least five
223 stages/tissues). Of these, 569 genes consistently used the same dominant uORF across the measured
224 samples, significantly higher than the number expected under randomness (5 genes, 95% confidence
225 interval: 1-10) based on shuffling the TEs of uORFs 1,000 times (Fig. S14). This trend was also
226 observed in *D. simulans* and persisted under different thresholds for defining "translated uORFs" (Fig.
227 S14). Furthermore, the dominant uORFs showed a higher proportion of conserved uATGs than the
228 other translated uORFs (median proportion 82.5% versus 78.8%, $P < 0.001$) (Fig. 2D). Additionally,
229 TE fold-changes ($|\log_2$TE-FC$|$) between the two species were, on an average, 23.2% smaller for
230 dominant uORFs than for other conserved uORFs (Fig. 2E), suggesting that dominant uORFs are more
231 likely under stronger stabilizing selection. These findings suggest that, in genes with multiple uORFs,
232 the dominantly translated uORF may play a more important role in regulating CDS translation than the
233 other uORFs."
234
235 **uORFs and CDSs show correlated translation differences between *D. melanogaster* and *D.***
236 ***simulans***
237 To investigate the relationship between translational changes in uORFs and their downstream CDSs
238 across species, we analyzed the translation efficiency (TE) of uORFs and corresponding downstream
239 CDSs in *D. melanogaster* and *D. simulans*. Consistent with our simulations that the translation of a
240 uORF is tightly linked to that of downstream CDS, uORFs exhibited a significant positive correlation
241 with the TE of their downstream CDSs in all samples analyzed ($P < 0.001$, Spearman's correlation)
242 (Fig. 3A). We then compared the interspecific TE change of a uORF ($\beta_u = TE_{uORF,sim}/TE_{uORF,mel}$)
243 with that of its corresponding CDS ($\beta_C = TE_{CDS,sim}/TE_{CDS,mel}$) between *D. melanogaster* and *D.*
244 *simulans*. We found β_u is significantly positively correlated with β_C across all samples (P values <
245 0.001, Spearman's correlation) (Fig. 3B). These results align well with our simulations, which showed
246 that fluctuations in translational factors (such as ribosomes) influence both uORF and CDS translation
247 in the same direction (Fig. S8-S9).
248
249 **uORFs buffer interspecific translational divergence of CDSs**
250 While the direction of translational efficiency (TE) changes for uORFs and CDSs tends to be consistent,
251 our simulations suggest that the magnitude of TE changes in CDSs is generally smaller due to the

252 buffering effect of uORF translation (Fig. 1B, S4-5). To validate this, we first identified uORFs and
253 CDSs with significant interspecific TE differences by assessing whether $\beta_u$or β_C significantly deviated
254 from 1, using an established statistical framework [23]. This analysis uncovered 1,151 to 4,189 CDSs with
255 significant interspecific TE changes (FDR < 0.05) (Table 1), with genes involved in development,
256 morphogenesis, and differentiation being significantly enriched during embryonic stages (Fig. S15).
257 Conversely, genes related to metabolism, response to stimuli, and signaling were enriched in larval,
258 pupal, and adult stages (Fig. S15). Additionally, we identified 144 to 1,193 uORFs with significant TE
259 differences between species, accounting for approximately 1-15% of expressed uORFs (Table 1). The
260 smaller number of uORFs showing significant TE changes compared to CDSs between *D. melanogaster*
261 and *D. simulans* likely reflects their shorter length and reduced statistical power, rather than indicating
262 that uORFs are less variable in translation than CDSs.
263
264 To further investigate, we quantitatively compared the magnitude of interspecific TE changes between
265 uORFs and their corresponding CDSs using a previous method [23]. We defined $\gamma = \beta_c/\beta_u$ for a uORF-
266 CDS pair within the same mRNA and tested whether γ was significantly different from 1 to identify
267 pairs with differential TE changes between uORFs and CDSs (Fig. S16). When γ < 1 and β_u >1, or γ >
268 1 and β_u <1, it indicates that the TE change for the CDS is smaller than that for the uORF (Fig. S16).
269 Among CDS-uORF pairs where β_u > 1, nearly all (8-487) showed a significant γ < 1 in each sample,
270 except for one pair from the pupal stage where γ > 1 (Table 1). This suggests that the magnitude of TE
271 changes in CDSs was generally smaller than in uORFs when uORF TE increased, and vice versa when
272 uORF TE decreased (β_u < 1) (Table 1). These comparative translatome analyses indicate that uORFs
273 buffered mRNA translation changes during the evolutionary divergence of *D. melanogaster* and *D.*
274 *simulans*.
275
276 **uORF buffering is influenced by its conservation, dominance, and length**
277 To investigate how the conservation level and translation patterns of uORFs influence their buffering
278 capacity on CDS translation, we categorized genes expressed in each pair of samples into three classes:
279 Class I, genes with uORFs conserved and dominantly translated in both *Drosophila* species; Class II,
280 genes with conserved uORFs translated in both species but not dominantly in at least one; and Class III,
281 the remaining expressed genes. We then compared the interspecific translation efficiency (TE) changes
282 of the CDS ($|\beta_c|$) across these three categories. Significant differences in $|\beta_c|$ were observed, with a
283 consistent hierarchy of Class I < II < III across all pairs of samples (Fig. 3C and S17). On average, Class
284 I genes exhibited an average of 8.18% and 23.8% lower $|\beta_c|$ values compared to Class II and Class III,
285 respectively. This indicates that conserved and dominantly translated uORFs exert a stronger buffering
286 effect on CDS translation.
287

288 To further validate the simulation results suggesting that longer uORFs have a stronger buffering effect
289 (Fig. 1C and S6-7), we divided genes expressed in each pair of samples into three groups: those without
290 translated uORFs (No), those with short uORFs (short, total length below the median), and those with
291 long uORFs (long, total length above the median). Consistently, longer uORFs were associated with
292 stronger buffering effects on CDS translation across all pairs of samples (Fig. 3D and S18). Specifically,
293 genes with longer uORFs showed 12.7% and 26.5% lower $|\beta_c|$ values compared to genes with short
294 uORFs or no uORFs, respectively.
295
296 Overall, these findings underscore that the buffering capability of a uORF is positively correlated with
297 its conservation level, translation dominance, and length.
298
299 **uORFs buffer translational fluctuations during *Drosophila* development**
300 Gene expression undergoes dynamic changes during *Drosophila* development, with significant
301 alterations in the translation program to meet developmental demands, including shifts in ribosome
302 loading rates [24,62,63]. Therefore, we extended our analysis to investigate the role of uORFs in buffering
303 these translational fluctuations, hypothesizing that uORFs could mitigate them during development.
304 According to our hypothesis, if a gene has a translated uORF in *D. melanogaster* but not in its
305 orthologous gene in *D. simulans*, then the translation of this gene is likely more stable across
306 developmental stages in *D. melanogaster* than its ortholog in *D. simulans*, and vice versa.
307
308 To test this hypothesis, we compared the CV of CDS TE across 10 developmental stages in *D.*
309 *melanogaster* and in *D. simulans*, respectively. Genes with translated uORFs (TE>0.1 in at least one
310 sample) in *D. melanogaster* exhibited significantly 22.5% smaller CVs in *D. melanogaster* than their
311 orthologs lacking these uORFs in *D. simulans*, indicating more stable translation (Fig. 4A). Consistently,
312 for genes with translated uORFs in *D. simulans* but not in *D. melanogaster*, the CVs were 13.3% lower
313 in *D. simulans* compared to their orthologs lacking these uORFs in *D. melanogaster* (Fig. 4B).
314 Consistent results were observed when a uORF is required to be translated (TE > 0.1) in all 10 samples
315 (Fig. S19A), in the 4 embryonic stages (Fig. S19B) or in the 6 stages including embryos, larva, and
316 pupa (Fig. S19C). Moreover, within each species, genes with translated uORFs also showed less
317 variability in CDS TE across developmental stages compared to those without translated uORFs, with
318 a 31.8% reduction in *D. melanogaster* and a 28.9% reduction in *D. simulans* (Fig. 4C). This effect was
319 consistent across different thresholds for defining "translated uORFs" (Fig. S20).
320
321 These results suggest that uORFs function as translational buffers, reducing gene translation
322 fluctuations during *Drosophila* development.
323
324 **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 ($|log_2(\beta_C)|$ < 0.05, adjusted P > 0.05), and 3) a significant difference between the magnitude of these changes ($|\log_2(\gamma)|$ > 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) analysis [18] 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.

361 ***bcd* uORF mutants show wide transcriptomic alteration during *Drosophila* embryogenesis**
362 Since Bcd regulates the expression of many zygotic genes [64-66], we anticipated that the increased
363 translation of *bcd* resulting from disrupting the uORF would influence *Drosophila* transcriptomes and
364 phenotypes. To verify this notion, we performed RNA sequencing on embryos from WT and
365 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
366 29°C conditions, using two biological replicates (Table S4 and Fig. S23). Differential expression
367 analysis revealed widespread alterations in gene expression between WT and mutant embryos, with the
368 number of differentially expressed genes (DEGs) increasing over developmental time and at higher
369 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
370 h embryos, respectively; while at 29°C, we detected 3,884, 2,358, 2,901, and 4,164 DEGs in the
371 corresponding stages (Fig. 6D). The majority of DEGs were stage-specific, with only a small fraction
372 consistently differentially expressed across all four stages. Functional enrichment analysis of the DEGs
373 revealed distinct biological pathways affected at each stage, including cell morphogenesis and pattern
374 specification in 0-2 h embryos, metabolic processes and tissue development in 2-6 h and 6-12 h embryos,
375 and mitochondrial respiration in 12-24 h embryos (Fig. 6E). Notably, direct targets of Bcd [72] were
376 significantly enriched among the DEGs in three out of the four stages, with the exception of 2-6 h
377 embryos (Fig. S24). RT-qPCR validation of 20 target genes of Bcd confirmed the reliability of the
378 RNA-seq differential expression analysis (Fig. S25). Together, these findings demonstrate that
379 disruption of the *bcd* uORF leads to widespread transcriptional changes during *Drosophila* development,
380 affecting processes ranging from embryogenesis to postembryonic metabolism.
381
382 ***bcd* uORF mutants display decreased hatching rates and starvation resistance**
383 Given the widespread transcriptome alterations, we anticipated phenotypic abnormalities in the *bcd*
384 uORF mutants. As expected, both uKO1/uKO1 and uKO2/uKO2 mutants exhibited significantly lower
385 hatching rates compared to WT ($P = 1.4\times10^{-6}$ and 7.9×10^{-6}, respectively, Wilcoxon rank sum test
386 [WRST]; Fig. 7A). At 25°C, uKO1/uKO1 mutants produced fewer offspring than WT flies ($P < 0.001$,
387 WRST, Fig. 7B). Given that *bcd* is a maternal gene, we expected reciprocal crosses between
388 uKO1/uKO1 mutants and WT flies to produce different outcomes. Indeed, crossing uKO1/uKO1 males
389 with WT females resulted in offspring numbers similar to those from WT crosses, while crossing
390 uKO1/uKO1 females with WT males yielded offspring numbers comparable to crosses between
391 uKO1/uKO1 mutants (Fig. 7B). This verified that the reduction in offspring is due to maternal defects
392 in the uKO1/uKO1 mutants. Similar patterns were observed for uKO2/uKO2 mutants at 25°C (Fig. 7B).
393 Notably, the fecundity reduction in uORF-KO mutants was more pronounced at 29°C (Fig. 7C).
394 Furthermore, crossing uKO1/uKO1 and uKO2/uKO2 mutants produced significantly fewer progeny
395 than WT flies (Fig. 7C), ruling out genetic background or off-target effects. Collectively, these data
396 demonstrate that disrupting the *bcd* uORF significantly impaired hatchability and fertility.
397

398 We also found both uKO1/uKO1 and uKO2/uKO2 female mutants perished significantly faster than
399 WT flies under starvation conditions (Fig. 7D). Males showed similar tendencies, although the
400 difference was not statistically significant for uKO1/uKO1 mutants (Fig. 7D). These data suggest that
401 the knockout of *bcd* uORF diminished starvation resistance in adults, likely due to embryogenesis
402 abnormalities induced by the *bcd* uORF deletion, even in those that successfully developed to adulthood.
403
404 **Conservation of uORF-mediated translational buffering in primates**
405 To explore the generality of uORF-mediated translational buffering across evolutionary clades, we
406 analyzed previously published transcriptome and translatome data from three tissues (brain, liver, and
407 testis) in humans and macaques [73]. We identified 33,680 canonical uORFs in humans and 29,516 in
408 macaques, with 24,385 conserved between the two species. Despite the larger number of uORFs in
409 primates compared to *Drosophila* due to differences in genome size and gene number, the median TE
410 of conserved uORFs was 1.79 times that of non-conserved uORFs in humans, and the corresponding
411 ratio was 3.43 in macaques (Fig. 8A and Fig. S26). TEs of uORFs were positively correlated between
412 humans and macaques across all tissues (P < 0.001, Fig. 8B). Additionally, significant positive
413 correlations were observed between the TEs of uORFs and their corresponding coding sequences (CDSs)
414 in all tissues (Fig. S27). Although interspecific TE divergence of uORFs (β_u) and CDSs (β_C) were
415 positively correlated (P < 0.001, Fig. 8C), uORFs generally exhibited larger divergence (Table S5).
416 Notably, longer uORFs showed stronger buffering effects on CDS translation, reducing interspecific
417 TE divergence by 12.9% and 38.0% compared to genes with short or no uORFs (Fig. 8D and Fig. S28).
418
419 As in *Drosophila*, we categorized expressed human genes into three classes based on uORF
420 conservation and translation: Class I, genes with conserved and dominantly translated uORFs in both
421 humans and macaques; Class II, genes with conserved uORFs translated in both species but not
422 dominantly in at least one; and Class III, the remaining expressed genes (Fig. 8E and Fig. S29).
423 Consistent with findings in *Drosophila*, significant differences in $|\beta_c|$ between humans and macaques
424 were observed in the order of Class I < II < III, with Class I genes showing 9.8% and 17.1% lower $|\beta_c|$
425 values compared to Class II and Class III genes, respectively (Fig. 8E and Fig. S29). These similarities
426 between primates and *Drosophila*—two clades that diverged over 700 million years ago [74] —suggest
427 that uORF-mediated translational buffering is a widespread mechanism for stabilizing gene translation
428 across evolutionary clades.
429
430 We also analyzed matched mRNA-Seq and Ribo-Seq data from 69 human lymphoblastoid cell lines
431 [75,76] to test whether uORFs buffer against translational variability across different individuals within
432 humans. Genes with translated uORFs exhibited, on average, 8.65% lower CVs in CDS TE across
433 individuals than genes without translated uORFs (Fig. 8F), with longer uORFs showing stronger

434 buffering effects (Fig. S30). Collectively, these findings suggest that uORFs play a crucial role in
435 reducing translational variability both across evolutionary clades and within species.
436
437 **Discussion**
438 Translational control is vital for maintaining protein homeostasis and cellular activities [77,78]. However,
439 the mechanisms underlying the conservation of protein abundance across species remain largely
440 unknown [5,6]. In this study, we extended the ICIER model and conducted simulations to explore uORFs'
441 regulatory roles in buffering translation during evolution. Our simulations demonstrated that uORF
442 translation reduces variability in downstream CDS translation, with buffering capacity positively
443 correlated with uORF translation efficiency, length, and number. Comparative translatome analyses
444 across developmental stages of two *Drosophila* species provided evidence that uORFs mitigate
445 interspecific differences in CDS translation. Similar patterns were observed between humans and
446 macaques, and between human individuals, suggesting that uORF-mediated buffering is an
447 evolutionarily conserved mechanism in animals. Additionally, in vivo experiments showed that
448 knocking out the *bcd* uORF led to aberrant embryogenesis and altered starvation resistance, with more
449 pronounced phenotypic effects at higher temperatures, supporting the role of uORFs in fine-tuning
450 translation and phenotypic outcomes.
451
452 While the prevailing consensus on uORFs' functions is that they repress the translation of downstream
453 CDS by sequestering ribosomes [34,43-48], recent studies based on suggest uORFs might play essential
454 roles in buffering translational noise and stabilizing protein expression [43,55,56]. While these studies have
455 primarily focused on the role of uORFs in single-gene cases or under stress conditions, our work
456 expands this knowledge by demonstrating that uORFs act as a general mechanism for buffering
457 translational variability on a genome-wide scale across multiple species and developmental stages. Our
458 findings also suggest that uORF-mediated translational buffering is not merely a response to
459 environmental stress but an evolutionarily conserved mechanism integral to maintaining protein
460 homeostasis across species. Organisms can buffer phenotypic variation against environmental or
461 genotypic perturbations through "canalization" [79]. The discovery that uORFs are crucial for *Drosophila*
462 development links them to canalization, providing evidence that uORFs reduce interspecific CDS
463 translation differences and enhance phenotypic stability. While previous studies primarily focused on
464 the stabilization of transcriptional level [80-84] or protein level [55,82,85-87], our study opens new avenues for
465 exploring how organisms maintain phenotypic robustness at the translational level through molecular
466 mechanisms like uORF-mediated buffering. This study lays the groundwork for future research into the
467 evolutionary and functional roles of uORFs, particularly in the context of canalization and adaptive
468 evolution.
469

470 The original ICIER model was devised to elucidate how a single uORF can confer resistance to global
471 translation inhibition on its host gene during stress conditions [43]. It proposed that an elongating ribosome
472 (80S) causes downstream 40S subunits to dissociate from the mRNA after collision. Our revised model
473 expands this scope to include bidirectional dissociation, where an 80S ribosome can release both
474 upstream and downstream 40S subunits, reflecting recent findings that 80S primarily causes upstream
475 40S dissociation during 40S/80S collisions [56]. Mechanistically, increasing the loading rate of 40S
476 ribosomes onto the mRNA (R_{in}) elevates the density of 40S and 80S ribosomes scanning the uORF,
477 thereby raising the likelihood of 40S-80S collisions and subsequent 40S dissociation. This leads to a
478 reduced flow of 40S ribosomes reaching the CDS relative to the initial increase in 40S at the 5' end of
479 the mRNA and uORF. Conversely, a decrease in 40S ribosome loading rate reduces the density of
480 scanning ribosomes, diminishing the probability of 40S-80S collisions and 40S dissociation, and
481 partially restoring the flow of 40S ribosomes to the CDS. Overall, the 40S-80S collision mechanism
482 and subsequent 40S dissociation at the uORF moderate downstream CDS translation less than changes
483 in ribosome loading at the 5' end of the mRNA or uORF. The revised model underscores the analogy
484 of uORFs as "molecular dams" adeptly modulating the stream of ribosomes and cushioning downstream
485 CDS translation against fluctuations. Further, the revised model integrates ribosomal collisions and
486 separations within the CDS located downstream of the uORF, shifting the focus from merely tracking
487 40S escape rates from the uORF to directly quantifying completed translations by 80S ribosomes at the
488 terminus of both uORFs and CDSs. This methodology provides a more direct and quantitative
489 assessment of CDS translation efficiency, offering new insights into the buffering role of uORFs.
490 Additionally, our study broadens its perspective to account for the potential interactive effects of
491 multiple uORFs within a gene, recognizing the prevalent clustering of these regulatory elements and
492 their combinatorial influence on translational control. In essence, our study presents a nuanced and
493 comprehensive expansion of the ICIER model, paving the way for a deeper understanding of uORFs as
494 pivotal elements in the maintenance of protein synthesis stability under variable translational conditions.
495 However, we realize the existence of alternative mechanisms governing uORF function, such as those
496 found on traditional repression models [55] or ribosome queuing and re-initiation strategies [56]. As such,
497 further computational studies are needed to dissect and understand the full spectrum of uORF-mediated
498 translational regulation.
499
500 The *bcd* gene, crucial for *Drosophila* embryogenesis, has been well studied [65,66,88-92], but its uORF
501 remains underexplored. Our study showed that deleting the *bcd* uORF led to abnormal phenotypes and
502 reduced fitness, underscoring the importance of uORF-mediated translation. These effects were more
503 severe under heat stress (29℃) than at normal temperatures (25℃), suggesting uORFs play a vital role
504 in buffering phenotypic variation. Rescue experiments, typically used to confirm that phenotypic
505 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 *w1118* 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 translatome 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.

527 **Materials and Methods**

528 **Modeling the uORF-mediated buffering effect on CDS translation**

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

540

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

555

556 The mRNA molecule was modeled as an array, and the value of each position in the array represented
557 the occupation status of ribosomes along the mRNA molecule. The process of translation was simulated
558 by a series of discrete actions referred to Andreev *et al.* [43]. Upon each action, there was a probability
559 that certain events would occur, including the addition of a 40S ribosome to the 5'-terminus, the
560 transformation of a 40S ribosome into an 80S ribosome at a CDS or uORF start codon, the movement
561 of a 40S ribosome or 80S ribosome to the next codon in the CDS or uORF, etc. More specific operations

562 in each action were described in Supplementary Materials and Methods. The simulation code is freely
563 available at GitHub: https://github.com/lujlab/uORF_buffer.
564
565 After 1,000,000 actions, we recorded the number of 80S ribosomes that completed translation at the
566 uORF (N_{Eu}) and CDS (N_{EC}) for each R_{in} input, which was used to represent the protein production
567 rates (i.e. translation efficiency) of the uORF and CDS, respectively. To test the effects of different
568 factors, we used various combinations of the parameters in Table S1 in the simulations. For a given
569 distribution of R_{in} input (1000 values) and the dissociation model, we obtained the corresponding 1000
570 N_{EC} values and their CV values by calculating the ratio of the standard deviation to the mean of N_{EC}.
571
572 In the double-uORF simulation, we only adopted a uniform distribution of R_{in} input and the
573 downstream dissociation model. Most parameters and simulation processes were similar to those in the
574 single-uORF simulation, except that an additional uORF was introduced upstream of the CDS with an
575 initiation probability of I_{uORF2}.
576
577 **Annotation of uORFs in *D. melanogaster* and *D. simulans***
578 We downloaded the whole-genome sequence alignment (maf) of *D. melanogaster* (dm6) and 26 other
579 insect species from the University of California Santa Cruz (UCSC) genome browser (genome.ucsc.edu)
580 [106]. To improve the genome correspondence between *D. melanogaster* and *D. simulans*, we adopted a
581 newly published reference-quality genome of *D. simulans* (NCBI, ASM438218v1) [107] to replace the
582 original *D. simulans* genome in UCSC maf. We softly masked repetitive sequences in the new genome
583 by RepeatMasker 4.1.1 (http://www.repeatmasker.org) and aligned it to dm6 with lastz [108] in
584 runLastzChain.sh following UCSC guidelines. The lastz alignment parameters and the scoring matrix
585 were the same as the parameters of dm6 and droSim1 in UCSC. Chained alignments were processed
586 into nets by the chainNet and netSyntenic programs. The alignment was integrated into the multiple
587 alignments of 27 species with multiz [109].
588
589 Based on the genome annotation of *D. melanogaster* (FlyBase r6.04, https://flybase.org/), we used the
590 Galaxy platform to parse the multiple sequence alignments of 5'UTRs in *Drosophila*. The 5'UTR
591 sequences of each annotated transcript from *D. melanogaster* and the corresponding sequences in *D.*
592 *simulans* were extracted from the maf. The start codons of putative uORFs were identified by scanning
593 all the ATG triplets (uATGs) within the 5'UTRs of *D. melanogaster* and *D. simulans*. uATGs that
594 overlapped with any annotated CDS region were removed. The presence or absence of each uATG of
595 *D. melanogaster* was determined at orthologous sites in *D. simulans* based on multiple genome
596 alignments, and vice versa. The conserved uORF was defined as a uORF where its uATG was present
597 in both *D. melanogaster* and the corresponding orthologous positions of *D. simulans*. For each protein-

598 coding gene, we only considered the canonical transcript. For transcripts containing multiple uORFs,
599 we defined the uORF that showed the highest TE as the dominant uORF. The branch length score (BLS)
600 of the uATGs was calculated as we previously described [18].
601
602 **Fly materials and general raising conditions**
603 The *sim4* strain of *D. simulans* was used to generate all the libraries of *D. simulans* in this study. All
604 flies were raised on standard corn medium and grown in 12 h light: 12 h dark cycles at 25℃ for general
605 conditions or at 29℃ for specific experimental design. The samples of embryos at different stages, larva,
606 pupa, bodies and heads were collected following a previous protocol [23].
607
608 **Processing *Drosophila* mRNA-Seq and Ribo-Seq data**
609 Ribo-Seq and matched mRNA-Seq libraries for different developmental stages and tissues of *D.*
610 *simulans* were constructed as we previously described [23] and sequenced on an Illumina HiSeq-2500
611 sequencer (run type: single-end; read length: 50 nt) according to the manufacturer's protocol. The 3'
612 adaptor sequences (TGGAATTCTCGGGTGCCAAGG) were trimmed using Cutadapt 3.0 with default
613 parameters [110], and the NGS reads were mapped to the genomes of yeast, *Wolbachia*, *Drosophila* viruses
614 and the sequences of tRNAs, ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) or small
615 nucleolar RNAs (snoRNAs) of *D. melanogaster* (FlyBase r6.04) and *D. simulans* (FlyBase r1.3) using
616 Bowtie2 version 2.2.3 [111] with the parameters -p8 --local -k1. The mapped reads of these
617 genomes/sequences were further removed in the downstream analysis.
618
619 After filtering, the mRNA-Seq and Ribo-Seq reads were mapped to the reference genomes of *D.*
620 *melanogaster* (FlyBase, r6.04) and *D. simulans* (NCBI, ASM438218v1) respectively using the Spliced
621 Transcripts Alignment to a Reference (STAR) algorithm [112]. For Ribo-Seq reads, we assigned a mapped
622 RPF (27-34 nt in length) to its P-site using the psite script from Plastid [113]. The uniquely mapped reads
623 were extracted and then mapped to the CDS of *D. melanogaster* and *D. simulans* respectively using
624 STAR [112]. The P-sites of RPF or mRNA-Seq reads that overlapped with a CDS were counted separately.
625 The reads that were not mapped to CDSs were then mapped to the 5'UTRs of *D. melanogaster* and *D.*
626 *simulans*. The P-sites of RPF or mRNA-Seq reads that overlapped with uORFs were counted separately.
627
628 The TE of a given uORF or CDS was calculated as the $RPKM_{P\text{-}site}/RPKM_{mRNA}$ ratio. For a few uORFs
629 with $RPKM_{mRNA} = 0$ but $RPKM_{P\text{-}site} > 0$, a pseudocount of 0.1 was added to both $RPKM_{mRNA}$ and
630 $RPKM_{P\text{-}site}$ to avoid dividing a positive value by zero. In each sample, expressed genes were defined as
631 the genes with a CDS $RPKM_{mRNA} > 0.1$ in both species, and translated uORFs were defined as uORFs
632 with a TE > 0.1.
633

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

648 $$\text{SE_log}_2(\beta_u) = \sqrt{(SE_log_2TE_{uORF,sim})^2 + (SE_log_2TE_{uORF,mel})^2}$$

649

650 As the Wald statistic, $\frac{log_2(\beta_u)}{\text{SE_log2}(\beta_u)}$, follows a standard normal distribution under the null hypothesis of
651 $\log_2(\beta_u) = 0$, we calculated the *P* value as follows: $2 \cdot (1 - \phi\,(|\frac{log_2(\beta_u)}{\text{SE_log2}(\beta_u)}|))$. The identification of
652 CDSs whose TEs differed significantly between the paired samples of *D. melanogaster* and *D.*
653 *simulans* was similar to the procedures above.
654 Then, we defined $\gamma = \frac{\beta_c}{\beta_u}$ (where $\beta_c = TE_{CDS,sim} / TE_{CDS,mel}$) and tested whether $\log_2(\gamma)$ was
655 significantly different from 0 to determine whether the magnitude of interspecific TE changes in CDSs
656 and uORFs was significantly different. We obtained $\log_2 TE_{uORF,sim}$, $\log_2 TE_{uORF,mel}$, $\log_2 TE_{CDS,sim}$,
657 and $\log_2 TE_{CDS,mel}$ as described above and estimated $SE_\log_2 TE_{uORF,sim}$, $SE_\log_2 TE_{uORF,mel}$,
658 $SE_\log_2 TE_{CDS,sim}$, and $SE_\log_2 TE_{CDS,mel}$ based on the biological replicates of female and male bodies
659 of *D. melanogaster*. Finally, $\log_2(\gamma)$ also follows a normal distribution with SE denoted as $\text{SE_log}_2(\gamma)$
660 =

661 $$\sqrt{(SE_log_2TE_{uORF,sim})^2 + (SE_log_2TE_{uORF,mel})^2 + (SE_log_2TE_{CDS,sim})^2 + (SE_log_2TE_{CDS,mel})^2}$$

662 As the Wald statistic, $\frac{log_2(\gamma)}{\text{SE_log2}(\gamma)}$, follows a standard normal distribution under the null hypothesis that
663 $\log_2(\beta_u) = 0$, we calculated the *P* value as follows: $2 \cdot (1 - \phi\,(|\frac{log_2(\gamma)}{\text{SE_log2}(\gamma)}|))$.
664
665 **Knocking out a uORF with CRISPR-Cas9 technology in *D. melanogaster***

666 We searched for possible sgRNA target sites near the uATG start codon of a uORF using the Benchling
667 website (https://www.benchling.com/crispr/) to design optimal single guide RNA (sgRNA) sequences
668 with high specificity and low off-target effects. We then synthesized single-stranded complementary
669 DNAs (ssDNAs) and annealed them to obtain double-stranded DNA (dsDNA), which served as the
670 template for sgRNA expression. The template sequences of the sgRNA used for *bcd* uATG-KO are
671 listed in Table S6. The dsDNA was then ligated into the BbsI-digested pU6B vector. The pU6B-sgRNA
672 plasmid was purified and injected into the embryos of transgenic Cas9 flies collected within one hour
673 of laying at the Tsinghua Fly Center as described in Ni *et al.* [114]. The injected embryos were kept at
674 25°C and 60% humidity until adulthood (G0). The G0 adult flies that hatched from injected embryos
675 were individually crossed with other flies (*y sc v*) to increase the number of offspring. Then, the F1
676 progeny were crossed with flies carrying an appropriate balancer (*Dr, e/TM3, Sb*). After F2 spawning,
677 the F1 individuals were screened for mutations of interest by genotyping. The primers used for
678 genotyping are listed in Table S6. The F2 progeny whose parents showed positive genotyping results
679 were then screened for the *yellow*$^{-}$ gene to separate the chromosome carrying *nos-Cas9*. The screened
680 F2 males were crossed with the flies containing the same balancer as above. After F2 genotyping, the
681 progeny (F3) of positive F2 individuals showing the same mutation status were crossed individually to
682 generate homozygous mutants in the F4 generation. The original homozygous mutants were
683 sequentially outcrossed with w^{1118} flies for 9 generations to purify the genetic background (Fig. S22B).
684

685 **Ribosome fraction analysis by sucrose gradient fractionation and RT–qPCR**

686 The 0-2 h embryos obtained from w^{1118}, uKO1/uKO1, and uKO2/uKO2 mutants raised at 25°C and
687 29°C were collected and homogenized in a Dounce homogenizer with lysis buffer [50 mM Tris pH 7.5,
688 150 mM NaCl, 5 mM $MgCl_2$, 1% Triton X-100, 2 mM dithiothreitol (DTT), 20 U/ml SuperaseIn
689 (Ambion), 0.5 tablets of proteinase inhibitor (Roche), 100 μg/ml emetine (Sigma Aldrich), and 50 μM
690 guanosine 5′-[β,γ-imido]triphosphate trisodium salt hydrate (GMP-PNP) (Sigma Aldrich)] at 4°C. The
691 lysates were clarified by centrifugation at 4°C and 20,000×g for 8 min, and the supernatants were
692 transferred to new 1.5 ml tubes. 10-45% sucrose gradients were prepared in buffer (250 mM NaCl, 50
693 mM Tris pH 7.5, 15 mM $MgCl_2$, 0.5 mM DTT, 12 U/ml RNaseOUT, 0.5 tablets of protease inhibitor,
694 and 20 μg/ml emetine) using a Gradient Master (Biocomp Instruments) in ULTRA-CLEAR Thinwall
695 Tubes (Beckman Coulter). A sample volume of up to 500 μl was applied to the top of each gradient.
696 After ultracentrifugation with a Hitachi P40ST rotor at 35,000 × rpm for 3 h at 4°C, the monosome and
697 polysome fractions were collected, flash-frozen in liquid nitrogen, and stored at -80°C until further use.
698

699 The RNA in the monosome and polysome fractions was extracted separately using TRIzol reagent (Life
700 Technologies, Inc.) and chloroform (Beijing Chemical Works) following the manufacturer's
701 instructions and were reverse transcribed into cDNA using the PrimeScript™ II 1st Strand cDNA
702 Synthesis Kit (Takara). RT–qPCR analysis of *bcd* cDNA and its targets was performed using

703 PowerUp™ SYBR™ Green Master Mix (Thermo Fisher) following the manufacturer's instructions
704 with *rp49* as an internal control. The primer sequences employed for RT-qPCR are listed in Table S6.
705 For each sample, the ratio of *bcd* mRNA abundance in the polysome fraction to that in the monosome
706 fraction was calculated as the P-to-M ratio. Six biological replicates were performed for each sample.
707
708 **Dual-luciferase reporter assays**
709 The wild-type (WT) 5'UTR of *bcd* was cloned from cDNA by PCR, and uATG mutations were
710 introduced into 5'UTR using specific amplification primers (listed in Table S6). The WT and mutated
711 5'UTR sequences were ligated into a linearized reporter plasmid (psiCHECK-2 vector, Promega). The
712 whole sequence of all the plasmids was validated by Sanger sequencing.
713
714 *Drosophila* S2 cells were cultured in Schneider's Insect Medium (Sigma) plus 10% (by volume) heat-
715 inactivated fetal bovine serum, 100 U/ml penicillin and 100 μg/ml streptomycin (Thermo Fisher) at 25℃
716 without CO_2 for 24 h to reach $1–2\times10^6$cells/ml before further treatments. Plasmid transfection was
717 conducted with Lipofectamine 3000 (L3000001, Thermo Fisher) according to the supplier's protocol.
718 The *Renilla* luciferase activity associated with WT or uORF-mutated 5'UTRs was measured according
719 to the manual of the Dual-Luciferase Reporter Assay System (Promega) 32 h after transfection and was
720 normalized to the activity of firefly luciferase.
721
722 **mRNA-Seq in the embryos of mutant and WT flies**
723 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
724 25°C and 29°C and conducted mRNA-Seq. Library construction and sequencing with PE150 were
725 conducted by Annoroad on the Illumina Nova6000 platform. Two biological replicates were sequenced
726 for each sample. The clean data were mapped to the reference genome of *D. melanogaster* (FlyBase,
727 r6.04) using STAR [112]. Reads mapped to the exons of each gene were tabulated with htseq-count [115].
728 The differentially expressed genes were identified using DESeq2 [116]. The Gene Ontology (GO) analyses
729 were conducted using the "clusterProfiler" package in R [117].
730
731 **Measurement of embryo hatchability**
732 We collected embryos from WT and mutant flies and manually seeded them in vials containing standard
733 corn medium at a density of 30 embryos per vial and 20 vials per strain. The embryos were cultivated
734 under standard conditions (60% humidity, 12 h light: 12 h dark cycles at 25°C). We counted the number
735 of pupae in each vial after the completion of pupation and calculated the corresponding hatching rate =
736 $number_{pupa}$ / 30
737
738 **Quantification of offspring number per female fly**

739 Newly hatched virgins were picked out and allowed to mature for two days in separate vials. They were
740 then mated by placing one virgin female with three male flies for two days. After that, each female
741 parent was transferred to a new vial to count the offspring number produced in each 10-day period at
742 25°C and 29°C. All assays were performed with 20 females per genotype.
743
744 **Measurement of starvation resistance in adult flies**
745 We selected 3- to 5-day-old adult males and females and placed them in the starvation medium (1.5%
746 agar), with 10 flies per vial and 10 vials for both males and females from each strain. We made
747 observations every 6 h or 12 h to count the number of deaths under starvation conditions until all flies
748 had starved to death. The survival curves were plotted by the ggsurvplot package in R.
749
750 **mRNA-Seq and Ribo-Seq data analysis in primates**
751 The mRNA-Seq and Ribo-Seq data from the brains, livers, and testes of humans and macaques were
752 downloaded from reference [73] with accession number E-MTAB-7247 (ArrayExpress). The mRNA-Seq
753 and Ribo-Seq data of human lymphoblastoid cell lines (LCL) from Yoruba individuals were
754 downloaded from references [75,76] under accession numbers GSE61742 and E-GEUV-1. The matched
755 mRNA-Seq and Ribo-Seq libraries of 69 individuals were used.
756
757 The uORF annotation and downstream analysis procedures for the human and macaque data were
758 similar to those applied in *Drosophila* as described above. The differential analysis of translational
759 efficiency in humans and macaques was conducted by Xtail [118]. In each pair of human-macaque samples,
760 expressed genes were defined as the genes with a CDS $RPKM_{mRNA} > 0.1$ in both species. The translated
761 uORFs in a sample were defined as uORFs with a TE > 0.1. For the human cell line data, expressed
762 genes were defined as genes with a mean CDS $RPKM_{mRNA} > 0.1$ across the cell lines, and translated
763 uORFs were defined as uORFs with a mean TE > 0.1
764
765 **Acknowledgments**
766 We thank Drs. Wei Xie, Weiwei Zhai, and Xionglei He for constructive comments and suggestions.
767 This work was supported by grants from the Ministry of Science and Technology of the People's
768 Republic of China (2022YFE0132000), the Yunnan Provincial Science and Technology Project at
769 Southwest United Graduate School (202302A0370006), the National Natural Science Foundation of
770 China (32070597) and the Natural Science Foundation of Beijing (5212006). We thank the National
771 Center for Protein Sciences at Peking University for their technical assistance. Some of the analyses
772 were performed on the High-Performance Computing Platform of the Center for Life Sciences.
773
774 **Declaration of interests**
775

776 The authors declare no competing interests.

777

778 **Data Availability Statement**

779 All deep-sequencing data generated in this study, including single-ended mRNA-Seq and Ribo-Seq
780 data of 10 developmental stages and tissues of *Drosophila simulans* and paired-end mRNA-Seq data of
781 0-2 h, 2-6 h, 6-12 h, and 12-24 h *Drosophila melanogaster* embryos, were deposited in the China
782 National Genomics Data Center Genome Sequence Archive (GSA) under accession numbers
783 CRA003198, CRA007425, and CRA007426. The mRNA-Seq and Ribo-Seq data for the different
784 developmental stages and tissues of *Drosophila melanogaster* were published in our previous paper [23]
785 and were deposited in the Sequence Read Archive (SRA) under accession number SRP067542.

786

787 **References**

788 1. Buccitelli, C., and Selbach, M. (2020). mRNAs, proteins and the emerging principles
789 of gene expression control. Nature Reviews Genetics *21*, 630-644. 10.1038/s41576-
790 020-0258-4.
791 2. Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in
792 disease. Cell *152*, 1237-1251. 10.1016/j.cell.2013.02.014.
793 3. MacNeil, L.T., and Walhout, A.J.M. (2011). Gene regulatory networks and the role of
794 robustness and stochasticity in the control of gene expression. Genome Research *21*,
795 645-657. 10.1101/gr.097378.109.
796 4. Hill, M.S., Vande Zande, P., and Wittkopp, P.J. (2021). Molecular and evolutionary
797 processes generating variation in gene expression. Nat Rev Genet *22*, 203-215.
798 10.1038/s41576-020-00304-w.
799 5. Vogel, C. (2013). Protein Expression Under Pressure. Science *342*, 1052-1053.
800 10.1126/science.1247833.
801 6. Signor, S.A., and Nuzhdin, S.V. (2018). The Evolution of Gene Expression in cis and
802 trans. Trends Genet *34*, 532-544. 10.1016/j.tig.2018.03.007.
803 7. Artieri, C.G., and Fraser, H.B. (2014). Evolution at two levels of gene expression in
804 yeast. Genome Res *24*, 411-421. 10.1101/gr.165522.113.
805 8. McManus, C.J., May, G.E., Spealman, P., and Shteyman, A. (2014). Ribosome
806 profiling reveals post-transcriptional buffering of divergent gene expression in yeast.
807 Genome Res *24*, 422-430. 10.1101/gr.164996.113.
808 9. Wang, Z., Sun, X., Zhao, Y., Guo, X., Jiang, H., Li, H., and Gu, Z. (2015). Evolution
809 of gene regulation during transcription and translation. Genome Biol Evol *7*, 1155-1167.
810 10.1093/gbe/evv059.
811 10. Khan, Z., Ford, M.J., Cusanovich, D.A., Mitrano, A., Pritchard, J.K., and Gilad, Y.
812 (2013). Primate transcript and protein expression levels evolve under compensatory
813 selection pressures. Science *342*, 1100-1104. 10.1126/science.1242379.
814 11. Wang, S.H., Hsiao, C.J., Khan, Z., and Pritchard, J.K. (2018). Post-translational
815 buffering leads to convergent protein expression levels between primates. Genome Biol
816 *19*, 83. 10.1186/s13059-018-1451-z.
817 12. Schrimpf, S.P., Weiss, M., Reiter, L., Ahrens, C.H., Jovanovic, M., Malmström, J.,
818 Brunner, E., Mohanty, S., Lercher, M.J., Hunziker, P.E., et al. (2009). Comparative
819 functional analysis of the Caenorhabditis elegans and Drosophila melanogaster
820 proteomes. PLoS Biol *7*, e48. 10.1371/journal.pbio.1000048.
821 13. Kusnadi, E.P., Timpone, C., Topisirovic, I., Larsson, O., and Furic, L. (2022).
822 Regulation of gene expression via translational buffering. Biochim Biophys Acta Mol
823 Cell Res *1869*, 119140. 10.1016/j.bbamcr.2021.119140.
824 14. Laurent, J.M., Vogel, C., Kwon, T., Craig, S.A., Boutz, D.R., Huse, H.K., Nozue, K.,
825 Walia, H., Whiteley, M., Ronald, P.C., and Marcotte, E.M. (2010). Protein abundances
826 are more conserved than mRNA abundances across diverse taxa. Proteomics *10*, 4209-
827 4212. 10.1002/pmic.201000327.
828 15. Teixeira, F.K., and Lehmann, R. (2019). Translational Control during Developmental
829 Transitions. Cold Spring Harb Perspect Biol *11*. 10.1101/cshperspect.a032987.
830 16. Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic
831 translation initiation and principles of its regulation. Nat Rev Mol Cell Biol *11*, 113-
832 127. 10.1038/nrm2838.
833 17. Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in
834 eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.
835 10.1016/j.cell.2009.01.042.

836 18. Zhang, H., Wang, Y., Wu, X., Tang, X., Wu, C., and Lu, J. (2021). Determinants of
837 genome-wide distribution and evolution of uORFs in eukaryotes. Nature
838 Communications *12*, 1076. 10.1038/s41467-021-21394-y.
839 19. Zhang, H., Wang, Y., and Lu, J. (2019). Function and Evolution of Upstream ORFs in
840 Eukaryotes. Trends Biochem Sci *44*, 782-794. 10.1016/j.tibs.2019.03.002.
841 20. Neafsey, D.E., and Galagan, J.E. (2007). Dual modes of natural selection on upstream
842 open reading frames. Mol Biol Evol *24*, 1744-1751. 10.1093/molbev/msm093.
843 21. Churbanov, A., Rogozin, I.B., Babenko, V.N., Ali, H., and Koonin, E.V. (2005).
844 Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs
845 of eukaryotic genes. Nucleic Acids Res *33*, 5512-5520. 10.1093/nar/gki847.
846 22. Resch, A.M., Ogurtsov, A.Y., Rogozin, I.B., Shabalina, S.A., and Koonin, E.V. (2009).
847 Evolution of alternative and constitutive regions of mammalian 5'UTRs. BMC
848 Genomics *10*, 162. 10.1186/1471-2164-10-162.
849 23. Zhang, H., Dou, S., He, F., Luo, J., Wei, L., and Lu, J. (2018). Genome-wide maps of
850 ribosomal occupancy provide insights into adaptive evolution and regulatory roles of
851 uORFs during Drosophila development. PLoS biology *16*, e2003903.
852 24. Malzer, E., Szajewska-Skuta, M., Dalton, L.E., Thomas, S.E., Hu, N., Skaer, H., Lomas,
853 D.A., Crowther, D.C., and Marciniak, S.J. (2013). Coordinate regulation of eIF2alpha
854 phosphorylation by PPP1R15 and GCN2 is required during Drosophila development. J
855 Cell Sci *126*, 1406-1415. 10.1242/jcs.117614.
856 25. Komonyi, O., Papai, G., Enunlu, I., Muratoglu, S., Pankotai, T., Kopitova, D., Maroy,
857 P., Udvardy, A., and Boros, I. (2005). DTL, the Drosophila homolog of PIMT/Tgs1
858 nuclear receptor coactivator-interacting protein/RNA methyltransferase, has an
859 essential role in development. J Biol Chem *280*, 12397-12404.
860 10.1074/jbc.M409251200.
861 26. Medenbach, J., Seiler, M., and Hentze, M.W. (2011). Translational control via protein-
862 regulated upstream open reading frames. Cell *145*, 902-913. 10.1016/j.cell.2011.05.005.
863 27. Chen, J., Tresenrider, A., Chia, M., McSwiggen, D.T., Spedale, G., Jorgensen, V., Liao,
864 H., van Werven, F.J., and Unal, E. (2017). Kinetochore inactivation by expression of a
865 repressive mRNA. Elife *6*. 10.7554/eLife.27417.
866 28. Cheng, Z., Otto, G.M., Powers, E.N., Keskin, A., Mertins, P., Carr, S.A., Jovanovic,
867 M., and Brar, G.A. (2018). Pervasive, Coordinated Protein-Level Changes Driven by
868 Transcript Isoform Switching during Meiosis. Cell *172*, 910-923 e916.
869 10.1016/j.cell.2018.01.035.
870 29. Kurihara, Y., Makita, Y., Kawashima, M., Fujita, T., Iwasaki, S., and Matsui, M. (2018).
871 Transcripts from downstream alternative transcription start sites evade uORF-mediated
872 inhibition of gene expression in Arabidopsis. Proc Natl Acad Sci U S A *115*, 7831-
873 7836. 10.1073/pnas.1804971115.
874 30. Yang, Y.F., Zhang, X., Ma, X., Zhao, T., Sun, Q., Huan, Q., Wu, S., Du, Z., and Qian,
875 W. (2017). Trans-splicing enhances translational efficiency in C. elegans. Genome Res
876 *27*, 1525-1535. 10.1101/gr.202150.115.
877 31. Wiestner, A., Schlemper, R.J., van der Maas, A.P., and Skoda, R.C. (1998). An
878 activating splice donor mutation in the thrombopoietin gene causes hereditary
879 thrombocythaemia. Nat Genet *18*, 49-52. 10.1038/ng0198-49.
880 32. Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N., and Hogg, D.
881 (1999). Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and
882 predisposes to melanoma. Nat Genet *21*, 128-132. 10.1038/5082.
883 33. Wen, Y., Liu, Y., Xu, Y., Zhao, Y., Hua, R., Wang, K., Sun, M., Li, Y., Yang, S., Zhang,
884 X.J., et al. (2009). Loss-of-function mutations of an inhibitory upstream ORF in the

885 human hairless transcript cause Marie Unna hereditary hypotrichosis. Nat Genet *41*,
886 228-233. 10.1038/ng.276.
887 34. Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames
888 cause widespread reduction of protein expression and are polymorphic among humans.
889 Proc Natl Acad Sci U S A *106*, 7507-7512. 10.1073/pnas.0810916106.
890 35. Lee, D.S.M., Park, J., Kromer, A., Baras, A., Rader, D.J., Ritchie, M.D., Ghanem, L.R.,
891 and Barash, Y. (2021). Disrupting upstream translation in mRNAs is associated with
892 human disease. Nat Commun *12*, 1515. 10.1038/s41467-021-21812-1.
893 36. Costa-Mattioli, M., and Walter, P. (2020). The integrated stress response: From
894 mechanism to disease. Science *368*. 10.1126/science.aat5314.
895 37. Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid
896 control of yeast. Annu Rev Microbiol *59*, 407-450.
897 10.1146/annurev.micro.59.031805.133833.
898 38. Lu, P.D., Harding, H.P., and Ron, D. (2004). Translation reinitiation at alternative open
899 reading frames regulates gene expression in an integrated stress response. J Cell Biol
900 *167*, 27-33. 10.1083/jcb.200408003.
901 39. Bohlen, J., Harbrecht, L., Blanco, S., Clemm von Hohenberg, K., Fenzl, K., Kramer,
902 G., Bukau, B., and Teleman, A.A. (2020). DENR promotes translation reinitiation via
903 ribosome recycling to drive expression of oncogenes including ATF4. Nat Commun *11*,
904 4676. 10.1038/s41467-020-18452-2.
905 40. Vasudevan, D., Neuman, S.D., Yang, A., Lough, L., Brown, B., Bashirullah, A.,
906 Cardozo, T., and Ryoo, H.D. (2020). Translational induction of ATF4 during integrated
907 stress response requires noncanonical initiation factors eIF2D and DENR. Nat
908 Commun *11*, 4677. 10.1038/s41467-020-18453-1.
909 41. Vattem, K.M., and Wek, R.C. (2004). Reinitiation involving upstream ORFs regulates
910 ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci U S A *101*, 11269-
911 11274. 10.1073/pnas.0400541101.
912 42. Andreev, D.E., O'Connor, P.B., Fahey, C., Kenny, E.M., Terenin, I.M., Dmitriev, S.E.,
913 Cormican, P., Morris, D.W., Shatsky, I.N., and Baranov, P.V. (2015). Translation of 5'
914 leaders is pervasive in genes resistant to eIF2 repression. Elife *4*, e03971.
915 10.7554/eLife.03971.
916 43. Andreev, D.E., Arnold, M., Kiniry, S.J., Loughran, G., Michel, A.M., Rachinskii, D.,
917 and Baranov, P.V. (2018). TASEP modelling provides a parsimonious explanation for
918 the ability of a single uORF to derepress translation during the integrated stress
919 response. Elife *7*. 10.7554/eLife.32563.
920 44. Kozak, M. (1989). The scanning model for translation: an update. J Cell Biol *108*, 229-
921 241. 10.1083/jcb.108.2.229.
922 45. Johnstone, T.G., Bazzini, A.A., and Giraldez, A.J. (2016). Upstream ORFs are
923 prevalent translational repressors in vertebrates. EMBO J *35*, 706-723.
924 10.15252/embj.201592759.
925 46. Hinnebusch, A.G., Ivanov, I.P., and Sonenberg, N. (2016). Translational control by 5'-
926 untranslated regions of eukaryotic mRNAs. Science *352*, 1413-1416.
927 10.1126/science.aad9868.
928 47. Morris, D.R., and Geballe, A.P. (2000). Upstream open reading frames as regulators of
929 mRNA translation. Mol Cell Biol *20*, 8635-8642. 10.1128/MCB.20.23.8635-
930 8642.2000.
931 48. Young, S.K., and Wek, R.C. (2016). Upstream Open Reading Frames Differentially
932 Regulate Gene-specific Translation in the Integrated Stress Response. J Biol Chem *291*,
933 16927-16935. 10.1074/jbc.R116.733899.

934 49. Dever, T.E., Feng, L., Wek, R.C., Cigan, A.M., Donahue, T.F., and Hinnebusch, A.G.
935 (1992). Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates
936 gene-specific translational control of GCN4 in yeast. Cell *68*, 585-596. 10.1016/0092-
937 8674(92)90193-g.
938 50. Palam, L.R., Baird, T.D., and Wek, R.C. (2011). Phosphorylation of eIF2 facilitates
939 ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. J Biol
940 Chem *286*, 10939-10949. 10.1074/jbc.M110.216093.
941 51. Baird, T.D., Palam, L.R., Fusakio, M.E., Willy, J.A., Davis, C.M., McClintick, J.N.,
942 Anthony, T.G., and Wek, R.C. (2014). Selective mRNA translation during eIF2
943 phosphorylation induces expression of IBTKalpha. Mol Biol Cell *25*, 1686-1697.
944 10.1091/mbc.E14-02-0704.
945 52. Chen, Y.J., Tan, B.C., Cheng, Y.Y., Chen, J.S., and Lee, S.C. (2010). Differential
946 regulation of CHOP translation by phosphorylated eIF4E under stress conditions.
947 Nucleic Acids Res *38*, 764-777. 10.1093/nar/gkp1034.
948 53. Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J., and Eisen, M.B. (2004). Noise
949 minimization in eukaryotic gene expression. PLoS Biol *2*, e137.
950 10.1371/journal.pbio.0020137.
951 54. Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory
952 networks. Proc Natl Acad Sci U S A *98*, 8614-8619. 10.1073/pnas.151588598.
953 55. Wu, H.W., Fajiculay, E., Wu, J.F., Yan, C.S., Hsu, C.P., and Wu, S.H. (2022). Noise
954 reduction by upstream open reading frames. Nat Plants *8*, 474-480. 10.1038/s41477-
955 022-01136-8.
956 56. Bottorff, T.A., Park, H., Geballe, A.P., and Subramaniam, A.R. (2022). Translational
957 buffering by ribosome stalling in upstream open reading frames. PLoS Genet *18*,
958 e1010460. 10.1371/journal.pgen.1010460.
959 57. Ciandrini, L., Stansfield, I., and Romano, M.C. (2010). Role of the particle's stepping
960 cycle in an asymmetric exclusion process: a model of mRNA translation. Phys Rev E
961 Stat Nonlin Soft Matter Phys *81*, 051904. 10.1103/PhysRevE.81.051904.
962 58. Reuveni, S., Meilijson, I., Kupiec, M., Ruppin, E., and Tuller, T. (2011). Genome-scale
963 analysis of translation elongation with a ribosome flow model. PLoS Comput Biol *7*,
964 e1002127. 10.1371/journal.pcbi.1002127.
965 59. von der Haar, T. (2012). Mathematical and Computational Modelling of Ribosomal
966 Movement and Protein Synthesis: an overview. Comput Struct Biotechnol J *1*,
967 e201204002. 10.5936/csbj.201204002.
968 60. Zhao, Y.B., and Krishnan, J. (2014). mRNA translation and protein synthesis: an
969 analysis of different modelling methodologies and a new PBN based approach. BMC
970 Syst Biol *8*, 25. 10.1186/1752-0509-8-25.
971 61. Tamura, K., Subramanian, S., and Kumar, S. (2004). Temporal Patterns of Fruit Fly
972 (Drosophila) Evolution Revealed by Mutation Clocks. Molecular Biology and
973 Evolution *21*, 36-44. 10.1093/molbev/msg236.
974 62. Kronja, I., Yuan, B., Eichhorn, S.W., Dzeyk, K., Krijgsveld, J., Bartel, D.P., and Orr-
975 Weaver, T.L. (2014). Widespread changes in the posttranscriptional landscape at the
976 Drosophila oocyte-to-embryo transition. Cell Rep *7*, 1495-1508.
977 10.1016/j.celrep.2014.05.002.
978 63. Qin, X., Ahn, S., Speed, T.P., and Rubin, G.M. (2007). Global analyses of mRNA
979 translational control during early Drosophila embryogenesis. Genome Biol *8*, R63.
980 10.1186/gb-2007-8-4-r63.
981 64. Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and
982 Nusslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the
983 anterior pattern of the Drosophila embryo. EMBO J *7*, 1749-1756.

65. Driever, W., and Nusslein-Volhard, C. (1988). A gradient of bicoid protein in Drosophila embryos. Cell *54*, 83-93. 10.1016/0092-8674(88)90182-1.
66. Driever, W., and Nusslein-Volhard, C. (1988). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. Cell *54*, 95-104. 10.1016/0092-8674(88)90183-3.
67. Abdelmohsen, K., Panda, A.C., Kang, M.J., Guo, R., Kim, J., Grammatikakis, I., Yoon, J.H., Dudekula, D.B., Noh, J.H., Yang, X., et al. (2014). 7SL RNA represses p53 translation by competing with HuR. Nucleic Acids Res *42*, 10099-10111. 10.1093/nar/gku686.
68. Panda, A.C., Abdelmohsen, K., Martindale, J.L., Di Germanio, C., Yang, X., Grammatikakis, I., Noh, J.H., Zhang, Y., Lehrmann, E., Dudekula, D.B., et al. (2016). Novel RNA-binding activity of MYF5 enhances Ccnd1/Cyclin D1 mRNA translation during myogenesis. Nucleic Acids Res *44*, 2393-2408. 10.1093/nar/gkw023.
69. Panda, A.C., Martindale, J.L., and Gorospe, M. (2017). Polysome Fractionation to Analyze mRNA Distribution Profiles. Bio Protoc *7*. 10.21769/BioProtoc.2126.
70. Amourda, C., Chong, J., and Saunders, T.E. (2018). MicroRNAs buffer genetic variation at specific temperatures during embryonic development. bioRxiv, 444810. 10.1101/444810.
71. Lu, G., Zhao, Y., Chen, Q., Lin, P., Tang, T., Tang, Z., Liufu, Z., and Wu, C.-I. (2021). When development is constantly but weakly perturbed - Canalization by microRNAs. bioRxiv, 2021.2009.2004.458966. 10.1101/2021.09.04.458966.
72. Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L., et al. (2008). Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol *6*, e27. 10.1371/journal.pbio.0060027.
73. Wang, Z.Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., et al. (2020). Transcriptome and translatome co-evolution in mammals. Nature *588*, 642-647. 10.1038/s41586-020-2899-z.
74. Hedges, S.B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics *22*, 2971-2972. 10.1093/bioinformatics/btl505.
75. Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., and Gilad, Y. (2015). Genomic variation. Impact of regulatory variation from RNA to protein. Science *347*, 664-667. 10.1126/science.1260793.
76. Lappalainen, T., Sammeth, M., Friedlander, M.R., t Hoen, P.A., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506-511. 10.1038/nature12531.
77. Stein, K.C., and 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., and Qian, S.B. (2013). Less is more: improving proteostasis by translation slow down. Trends Biochem Sci *38*, 585-591. 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. 10.1038/150563a0.
80. Payne, J.L., and Wagner, A. (2015). Mechanisms of mutational robustness in transcriptional regulation. Front Genet *6*, 322. 10.3389/fgene.2015.00322.

1033 81. Denby, C.M., Im, J.H., Yu, R.C., Pesce, C.G., and Brem, R.B. (2012). Negative
1034 feedback confers mutational robustness in yeast transcription factor regulation. Proc
1035 Natl Acad Sci U S A *109*, 3874-3878. 10.1073/pnas.1116360109.
1036 82. Jarosz, D.F., Taipale, M., and Lindquist, S. (2010). Protein homeostasis and the
1037 phenotypic manifestation of genetic diversity: principles and mechanisms. Annu Rev
1038 Genet *44*, 189-216. 10.1146/annurev.genet.40.110405.090412.
1039 83. Ebert, M.S., and Sharp, P.A. (2012). Roles for microRNAs in conferring robustness to
1040 biological processes. Cell *149*, 515-524. 10.1016/j.cell.2012.04.005.
1041 84. Lu, G.A., Zhang, J., Zhao, Y., Chen, Q., Lin, P., Tang, T., Tang, Z., Wen, H., Liufu, Z.,
1042 and Wu, C.I. (2023). Canalization of Phenotypes-When the Transcriptome is
1043 Constantly but Weakly Perturbed. Mol Biol Evol *40*. 10.1093/molbev/msad005.
1044 85. Alon, U. (2007). Network motifs: theory and experimental approaches. Nat Rev Genet
1045 *8*, 450-461. 10.1038/nrg2102.
1046 86. Somogyvari, M., Khatatneh, S., and Soti, C. (2022). Hsp90: From Cellular to
1047 Organismal Proteostasis. Cells *11*. 10.3390/cells11162479.
1048 87. Zabinsky, R.A., Mason, G.A., Queitsch, C., and Jarosz, D.F. (2019). It's not magic -
1049 Hsp90 and its effects on genetic and epigenetic variation. Semin Cell Dev Biol *88*, 21-
1050 35. 10.1016/j.semcdb.2018.05.015.
1051 88. Cho, P.F., Poulin, F., Cho-Park, Y.A., Cho-Park, I.B., Chicoine, J.D., Lasko, P., and
1052 Sonenberg, N. (2005). A new paradigm for translational control: inhibition via 5'-3'
1053 mRNA tethering by Bicoid and the eIF4E cognate 4EHP. Cell *121*, 411-423.
1054 10.1016/j.cell.2005.02.024.
1055 89. Singh, A.P., Wu, P., Ryabichko, S., Raimundo, J., Swan, M., Wieschaus, E., Gregor,
1056 T., and Toettcher, J.E. (2022). Optogenetic control of the Bicoid morphogen reveals
1057 fast and slow modes of gap gene regulation. Cell Rep *38*, 110543.
1058 10.1016/j.celrep.2022.110543.
1059 90. Hannon, C.E., Blythe, S.A., and Wieschaus, E.F. (2017). Concentration dependent
1060 chromatin states induced by the bicoid morphogen gradient. Elife *6*.
1061 10.7554/eLife.28275.
1062 91. Struhl, G., Struhl, K., and Macdonald, P.M. (1989). The gradient morphogen bicoid is
1063 a concentration-dependent transcriptional activator. Cell *57*, 1259-1273. 10.1016/0092-
1064 8674(89)90062-7.
1065 92. Dubnau, J., and Struhl, G. (1996). RNA recognition and translational regulation by a
1066 homeodomain protein. Nature *379*, 694-699. 10.1038/379694a0.
1067 93. Wethmar, K., Begay, V., Smink, J.J., Zaragoza, K., Wiesenthal, V., Dorken, B.,
1068 Calkhoven, C.F., and Leutz, A. (2010). C/EBPbetaDeltauORF mice--a genetic model
1069 for uORF-mediated translational control in mammals. Genes Dev *24*, 15-20.
1070 10.1101/gad.557910.
1071 94. Miyake, T., Inoue, Y., Shao, X., Seta, T., Aoki, Y., Nguyen Pham, K.T., Shichino, Y.,
1072 Sasaki, J., Sasaki, T., Ikawa, M., et al. (2023). Minimal upstream open reading frame
1073 of Per2 mediates phase fitness of the circadian clock to day/night physiological body
1074 temperature rhythm. Cell Rep *42*, 112157. 10.1016/j.celrep.2023.112157.
1075 95. Xing, S., Chen, K., Zhu, H., Zhang, R., Zhang, H., Li, B., and Gao, C. (2020). Fine-
1076 tuning sugar content in strawberry. Genome Biol *21*, 230. 10.1186/s13059-020-02146-
1077 5.
1078 96. Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., Wang, D., and Gao, C. (2018).
1079 Genome editing of upstream open reading frames enables translational control in plants.
1080 Nat Biotechnol *36*, 894-898. 10.1038/nbt.4202.

1081 97. Si, X., Zhang, H., Wang, Y., Chen, K., and Gao, C. (2020). Manipulating gene
1082 translation in plants by CRISPR-Cas9-mediated genome editing of upstream open
1083 reading frames. Nat Protoc *15*, 338-363. 10.1038/s41596-019-0238-3.
1084 98. Xue, C., Qiu, F., Wang, Y., Li, B., Zhao, K.T., Chen, K., and Gao, C. (2023). Tuning
1085 plant phenotypes by precise, graded downregulation of gene expression. Nat
1086 Biotechnol. 10.1038/s41587-023-01707-w.
1087 99. Ivanov, I.P., Shin, B.S., Loughran, G., Tzani, I., Young-Baird, S.K., Cao, C., Atkins,
1088 J.F., and Dever, T.E. (2018). Polyamine Control of Translation Elongation Regulates
1089 Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. Mol Cell *70*,
1090 254-264 e256. 10.1016/j.molcel.2018.03.015.
1091 100. Luo, Z., and Sachs, M.S. (1996). Role of an upstream open reading frame in mediating
1092 arginine-specific translational control in Neurospora crassa. J Bacteriol *178*, 2172-2177.
1093 10.1128/jb.178.8.2172-2177.1996.
1094 101. Lovett, P.S., and Rogers, E.J. (1996). Ribosome regulation by the nascent peptide.
1095 Microbiol Rev *60*, 366-385. 10.1128/mr.60.2.366-385.1996.
1096 102. Vilela, C., and McCarthy, J.E. (2003). Regulation of fungal gene expression via short
1097 open reading frames in the mRNA 5'untranslated region. Mol Microbiol *49*, 859-867.
1098 10.1046/j.1365-2958.2003.03622.x.
1099 103. Raney, A., Law, G.L., Mize, G.J., and Morris, D.R. (2002). Regulated translation
1100 termination at the upstream open reading frame in s-adenosylmethionine decarboxylase
1101 mRNA. J Biol Chem *277*, 5988-5994. 10.1074/jbc.M108375200.
1102 104. Lin, Y., May, G.E., Kready, H., Nazzaro, L., Mao, M., Spealman, P., Creeger, Y., and
1103 McManus, C.J. (2019). Impacts of uORF codon identity and position on translation
1104 regulation. Nucleic Acids Res *47*, 9358-9367. 10.1093/nar/gkz681.
1105 105. Meijer, H.A., and Thomas, A.A. (2003). Ribosomes stalling on uORF1 in the Xenopus
1106 Cx41 5' UTR inhibit downstream translation initiation. Nucleic Acids Res *31*, 3174-
1107 3184. 10.1093/nar/gkg429.
1108 106. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M.,
1109 Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC
1110 Genome Browser database: 2015 update. Nucleic Acids Res *43*, D670-681.
1111 10.1093/nar/gku1177.
1112 107. Chakraborty, M., Chang, C.H., Khost, D.E., Vedanayagam, J., Adrion, J.R., Liao, Y.,
1113 Montooth, K.L., Meiklejohn, C.D., Larracuente, A.M., and Emerson, J.J. (2021).
1114 Evolution of genome structure in the Drosophila simulans species complex. Genome
1115 Res *31*, 380-396. 10.1101/gr.263442.120.
1116 108. Chiaromonte, F., Yap, V.B., and Miller, W. (2002). Scoring pairwise genomic sequence
1117 alignments. Pac Symp Biocomput, 115-126. 10.1142/9789812799623_0012.
1118 109. Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F., Roskin, K.M., Baertsch,
1119 R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Aligning multiple
1120 genomic sequences with the threaded blockset aligner. Genome Res *14*, 708-715.
1121 10.1101/gr.1933104.
1122 110. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput
1123 sequencing reads. EMBnet. journal *17*, 10-12.
1124 111. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2.
1125 Nat Methods *9*, 357-359. 10.1038/nmeth.1923.
1126 112. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,
1127 Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner.
1128 Bioinformatics *29*, 15-21.

113. Dunn, J.G., and Weissman, J.S. (2016). Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. BMC Genomics *17*, 958. 10.1186/s12864-016-3278-x.
114. Ni, J.Q., Zhou, R., Czech, B., Liu, L.P., Holderbaum, L., Yang-Zhou, D., Shim, H.S., Tao, R., Handler, D., Karpowicz, P., et al. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nat Methods *8*, 405-407. 10.1038/nmeth.1592.
115. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169. 10.1093/bioinformatics/btu638.
116. McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res *40*, 4288-4297. 10.1093/nar/gks042.
117. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284-287. 10.1089/omi.2011.0118.
118. Xiao, Z., Zou, Q., Liu, Y., and Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. Nat Commun *7*, 11194. 10.1038/ncomms11194.

1149 **Main Tables and Figures**

1150 **Table 1.** Numbers of genes showing different magnitudes of TE changes between uORFs and CDS at
1151 the interspecific level.

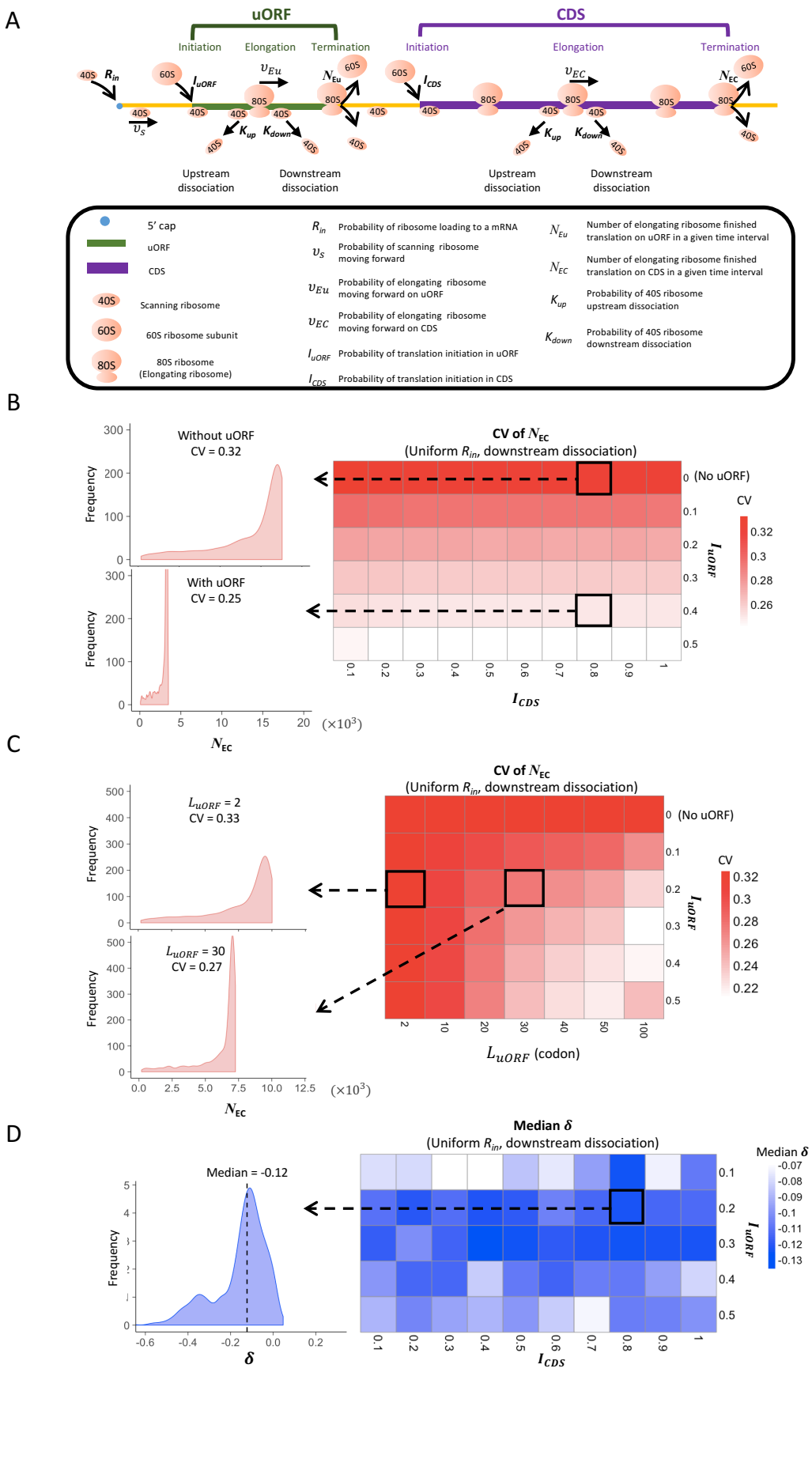
1152

Sample types	# of expressed uORFs *	$\beta_u \neq 1$ (%) **	# of expressed 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

1153 * Only uORFs and CDSs with an mRNA RPKM > 0.1 in both *D. melanogaster* and *D. simulans* were
1154 considered in each sample pair in the analysis.

1155 ** $\beta_u = TE_{uORF,sim}/TE_{uORF,mel}$ is the fold change of TE_{uORF} in *D. simulans* relative to *D. melanogaster*
1156 for each sample. $\beta_c = TE_{CDS,sim}/TE_{CDS,mel}$ is the fold change of TE_{CDS} in *D. simulans* relative to *D.*
1157 *melanogaster* for each sample. $\gamma = \beta_c/\beta_u$. For each CDS–uORF pair, either β_u>1 and γ < 1 or β_u<1
1158 and γ > 1 means that the magnitude of TE change is lower for a CDS than for a uORF. The statistical
1159 significance of β_c, β_u, and γ were determined according to an FDR < 0.05.

1160


A
uORF
CDS
Initiation
Elongation
Termination
Upstream dissociation
Downstream dissociation
5′ cap
uORF
CDS
Scanning ribosome
60S ribosome subunit
80S ribosome (Elongating ribosome)
Probability of ribosome loading to a mRNA
Probability of scanning ribosome moving forward
Probability of elongating ribosome moving forward on uORF
Probability of elongating ribosome moving forward on CDS
Probability of translation initiation in uORF
Probability of translation initiation in CDS
Number of elongating ribosome finished translation on uORF in a given time interval
Number of elongating ribosome finished translation on CDS in a given time interval
Probability of 40S ribosome upstream dissociation
Probability of 40S ribosome downstream dissociation
B
Without uORF
CV = 0.32
With uORF
CV = 0.25
Frequency
CV of N_{EC}
(Uniform R_{in}, downstream dissociation)
(No uORF)
CV
I_{uORF}
I_{CDS}
N_{EC}
(×10³)
C
L_{uORF} = 2
CV = 0.33
L_{uORF} = 30
CV = 0.27
L_{uORF} (codon)
D
Median = -0.12
Median δ
(Uniform R_{in}, downstream dissociation)
δ

1163 **Figure 1. Modeling simulation of uORF-mediated translation buffering.** (A) Model schema of the
1164 modified ICIER model (on the top); the parameters are listed in the box below the schema. (B) Heatmap
1165 showing the CVs of CDS TE (N_{EC}) under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with a
1166 uniform distribution of R_{in} input and the downstream dissociation model. The left panels elicited by the
1167 dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC}
1168 under I_{CDS} = 0.8 & I_{uORF} = 0 (top panel, without uORF) and I_{CDS} = 0.8 & I_{uORF} = 0.4 (bottom panel, with
1169 uORF). (C) Heatmap showing CVs of CDS TE (N_{EC}) under different L_{uORF} (x-axis) and I_{uORF} (y-axis)
1170 combinations with a uniform distribution of R_{in} input and the downstream dissociation model. The left
1171 panels elicited by the dotted lines from specific squares of right heatmap were two examples showing
1172 the distribution of N_{EC} under L_{uORF} = 2 & I_{uORF} = 0.2 (top panel) and L_{uORF} = 30 & I_{uORF} = 0.2 (bottom
1173 panel). (D) Heatmap showing median δ $[log_2(\Delta N_{EC}/\Delta N_{EU1})]$ under different I_{CDS} (x-axis) and I_{uORF}
1174 (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model.
1175 The left panel elicited by the dotted line from a specific square of right heatmap was an example
1176 showing the distribution of δ under I_{CDS} = 0.8 & I_{uORF} = 0.2. The vertical dashed line indicated the
1177 median value of δ.
1178

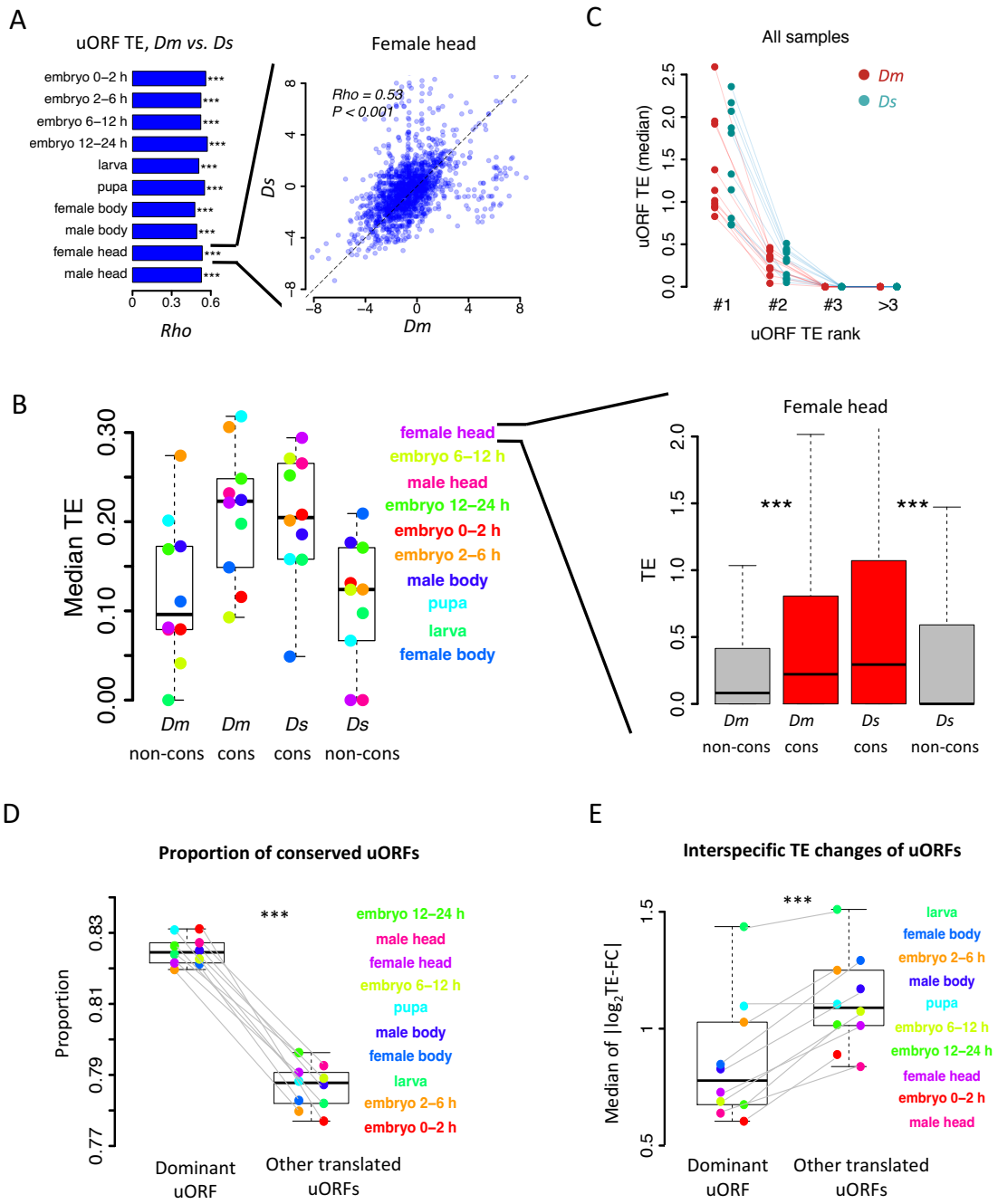


Figure 2. Conservation and translation of uORFs between *D. melanogaster* and *D. simulans*. (A) Spearman's correlation coefficients (*Rho*, represented by the bars) of conserved TEs between *Dm* (*D. melanogaster*) and *Ds* (*D. simulans*). ***, $P < 0.001$. Data for the female head sample is shown 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

1190 defined as the dominant uORF (#1). “#2” represents the second highest uORF TE and the same goes
1191 for “#3” and “>3”. Each dot represents the median TE of a sample. (D) Fraction of conserved uORFs
1192 among dominant uORFs and other translated uORFs in each sample. The paired samples in *Dm* and *Ds*
1193 were linked together. The *P* value was obtained by the paired Wilcoxon signed rank test. ***, $P < 0.001$.
1194 (E) Absolute values of the interspecific TE fold changes ($\log_2$TE-FC) of dominant uORFs and the other
1195 translated uORFs in each sample. The paired samples in *Dm* and *Ds* were linked together. The median
1196 value of each sample is shown. The *P* value was obtained via the paired Wilcoxon signed rank test. ***,
1197 $P < 0.001$. Data from the female head sample were used as an example in the right panel.
1198

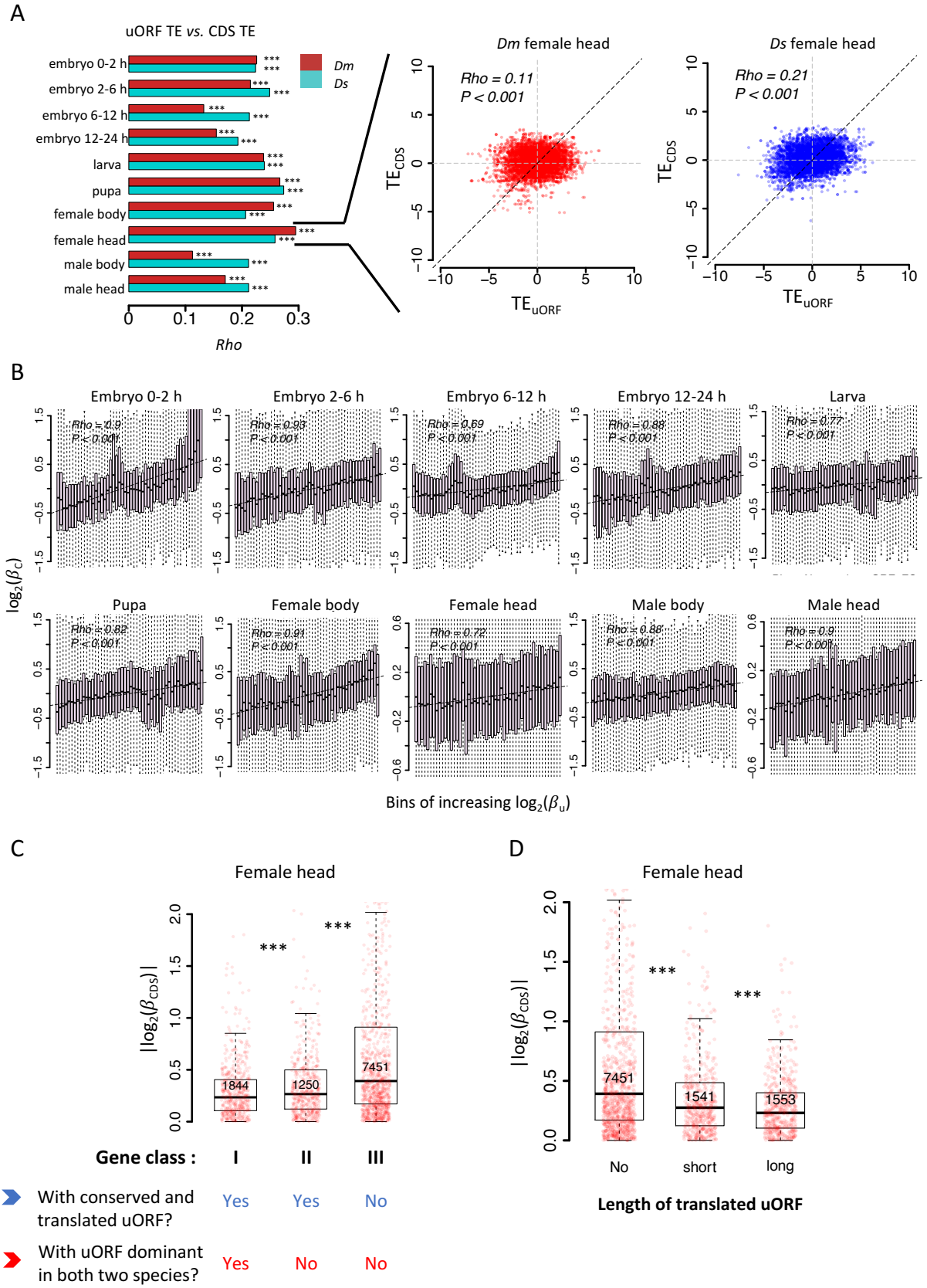
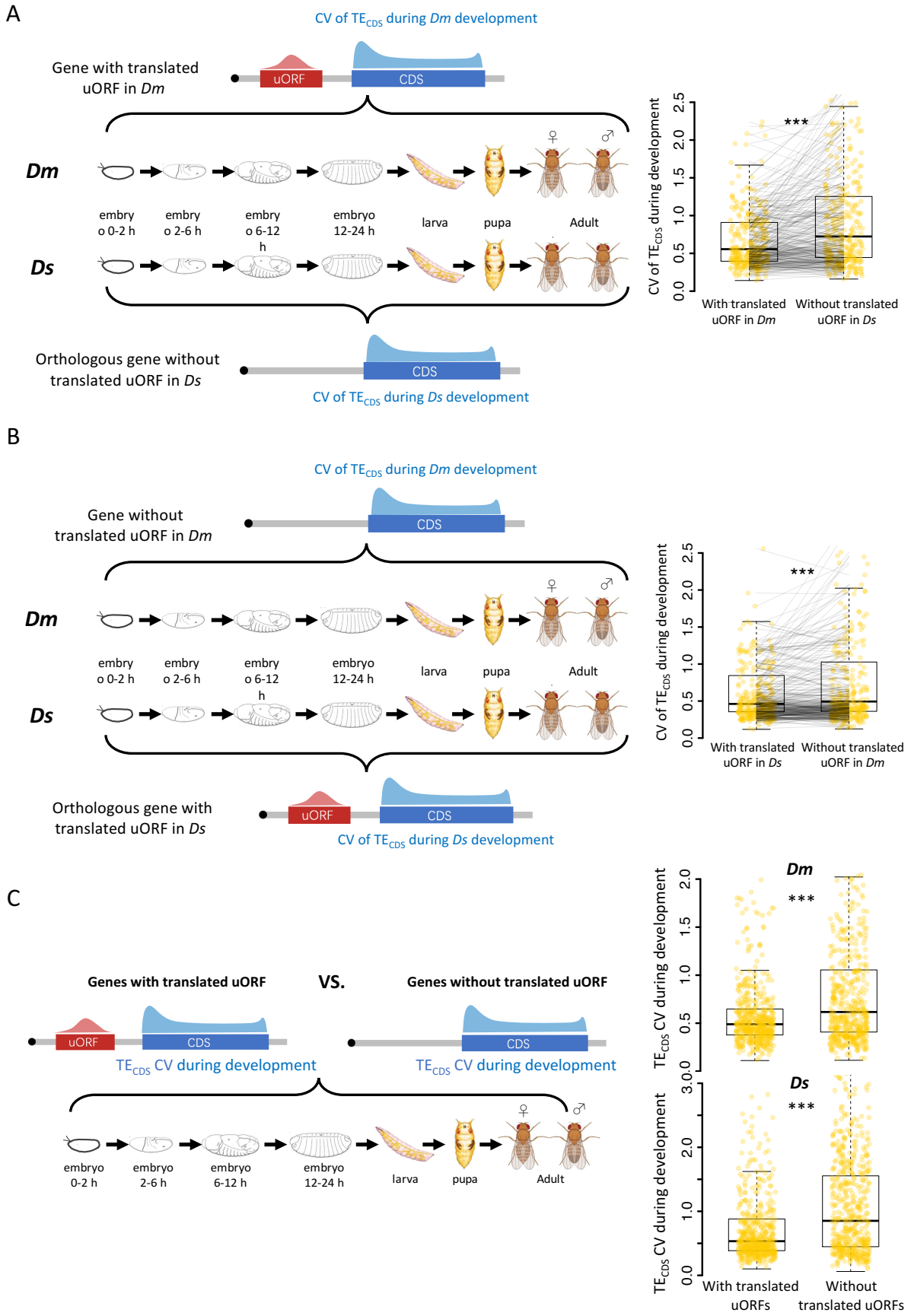


Figure 3. uORFs reduce CDS translational divergence between *D. melanogaster* and *D. simulans*. (A) The correlation of uORF TEs and the corresponding CDS TEs in 10 samples of *Dm* (*D.*

1203 *melanogaster*) and *Ds* (*D. simulans*). The bars represent Spearman's correlation coefficient (*Rho*). In
1204 all samples, we obtained both *P* values < 0.001. Data for the female head sample of *Dm* and *Ds* are
1205 shown as examples in the right panel. (B) Correlations between interspecific uORF TE changes ($\log_2\beta_u$)
1206 and CDS TE changes ($\log_2\beta_C$) in 10 samples. The x-axis was divided into 50 equal bins with increasing
1207 β_u. Spearman's correlation coefficients (*Rho*) are shown at the top left. ***, *P* < 0.001 in the correlation
1208 test. (C) Genes expressed in female heads (mRNA RPKM > 0.1 in both species) were classified into
1209 three classes according to whether a gene had a conserved and dominantly translated uORF or not.
1210 Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P* values were
1211 calculated using Wilcoxon rank sum tests between the neighboring groups. ***, *P* < 0.001. (D) Genes
1212 expressed in female heads were classified into three classes according to the length of translated uORFs.
1213 Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P* values were
1214 calculated using Wilcoxon rank sum tests between the neighboring groups. ***, *P* < 0.001.
1215


A
CV of TE$_{CDS}$ during *Dm* development
Gene with translated uORF in *Dm*
uORF
CDS
Dm
Ds
embryo 0-2 h
embryo 2-6 h
embryo 6-12 h
embryo 12-24 h
larva
pupa
Adult
Orthologous gene without translated uORF in *Ds*
CDS
CV of TE$_{CDS}$ during *Ds* development
CV of TE$_{CDS}$ during development
***
With translated uORF in *Dm*
Without translated uORF in *Ds*
B
CV of TE$_{CDS}$ during *Dm* development
Gene without translated uORF in *Dm*
CDS
Dm
Ds
embryo 0-2 h
embryo 2-6 h
embryo 6-12 h
embryo 12-24 h
larva
pupa
Adult
Orthologous gene with translated uORF in *Ds*
uORF
CDS
CV of TE$_{CDS}$ during *Ds* development
CV of TE$_{CDS}$ during development
***
With translated uORF in *Ds*
Without translated uORF in *Dm*
C
Genes with translated uORF
VS.
Genes without translated uORF
uORF
CDS
TE$_{CDS}$ CV during development
CDS
TE$_{CDS}$ CV during development
embryo 0-2 h
embryo 2-6 h
embryo 6-12 h
embryo 12-24 h
larva
pupa
Adult
Dm
***
Ds
***
TE$_{CDS}$ CV during development
With translated uORFs
Without translated uORFs

1216
1217

1218 **Figure 4. uORFs could reduce CDS translational fluctuation during *Drosophila* development.** (A)
1219 The CV of TE_{CDS} across 10 *Dm* (*D. melanogaster*) samples and 10 *Ds* (*D. simulans*) samples. The
1220 selected gene with uORFs translated (TE > 0.1) in at least one *Dm* sample but its homologous gene
1221 without translated uORF in *Ds* samples. Each pair of dots linked by a gray line represents a pair of
1222 homologous genes in *Dm* and *Ds*. ***, $P < 0.001$, Wilcoxon signed-rank test. (B) The CV of TE_{CDS}
1223 across 10 *Dm* samples and 10 *Ds* samples. The selected gene with uORFs translated (TE > 0.1) in at
1224 least one *Ds* sample but its homologous gene without translated uORF in *Dm* samples. Each pair of dots
1225 linked by a gray line represents a pair of homologous genes in *Dm* and *Ds*. ***, $P < 0.001$, Wilcoxon
1226 signed-rank test. (C) Within each *Drosophila* species, the CV of TE_{CDS} of genes with translated uORFs
1227 compared to genes without the translated uORFs. The *P* values are obtained by the Wilcoxon rank sum
1228 test. ***, $P < 0.001$.
1229

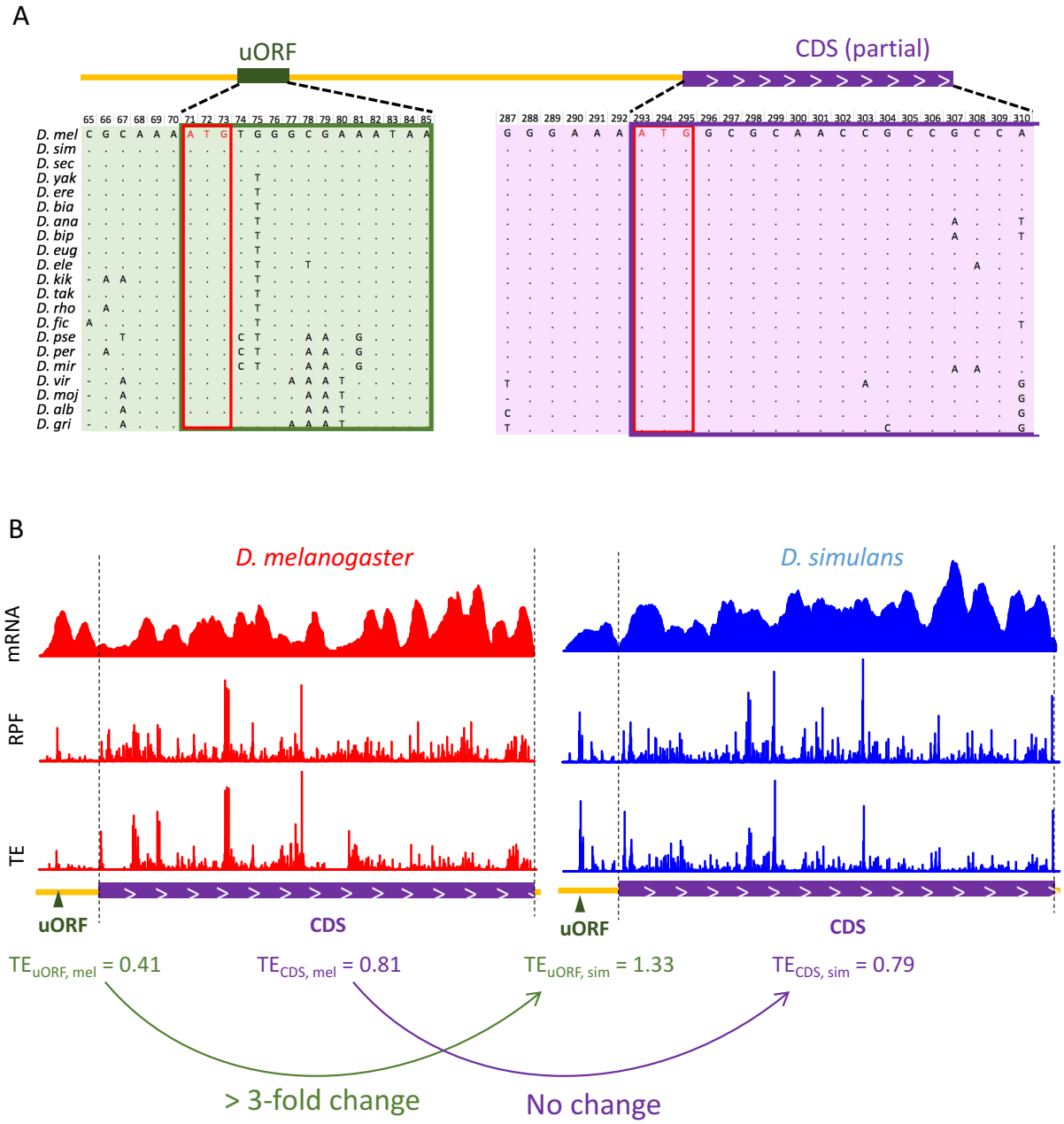
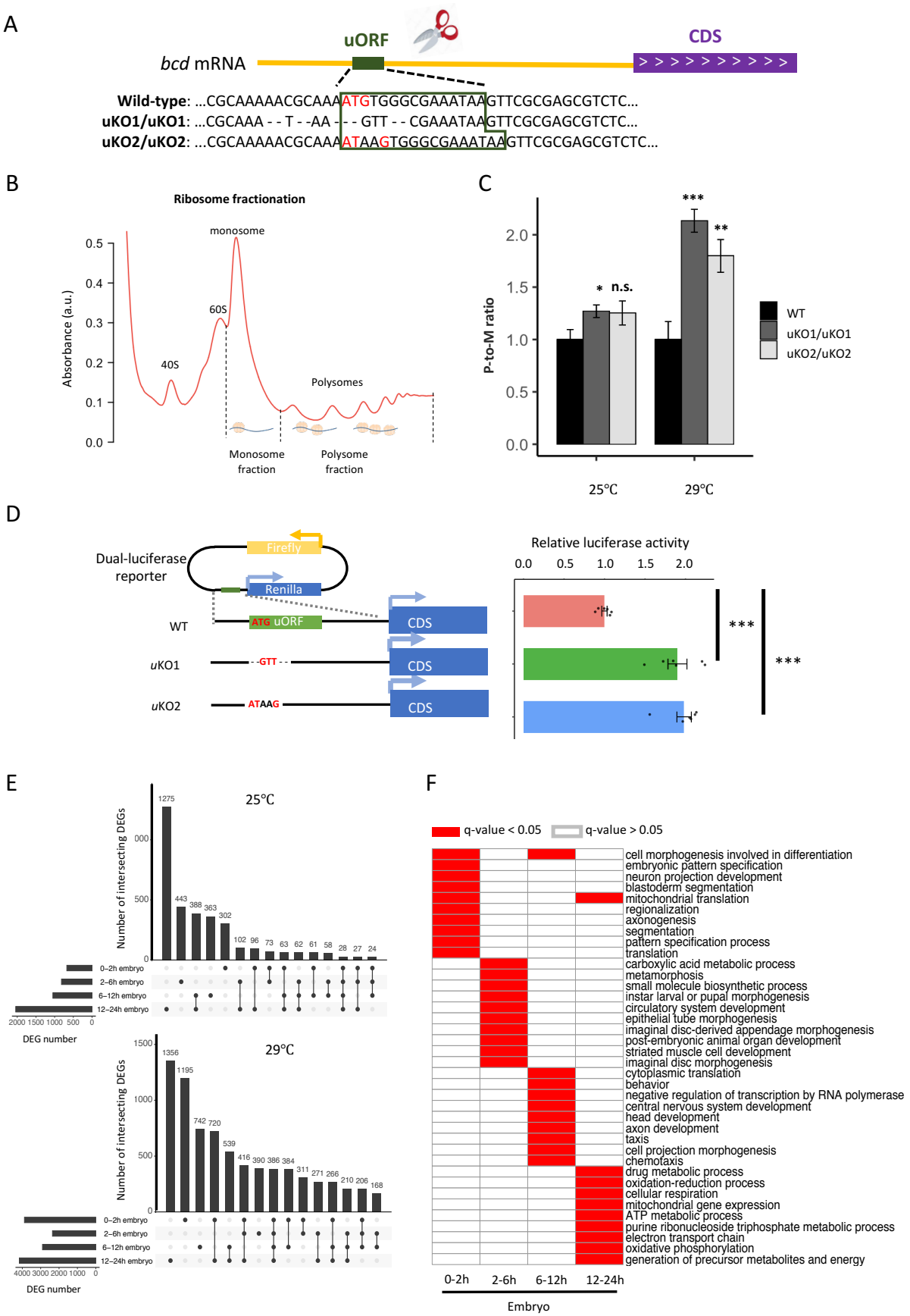
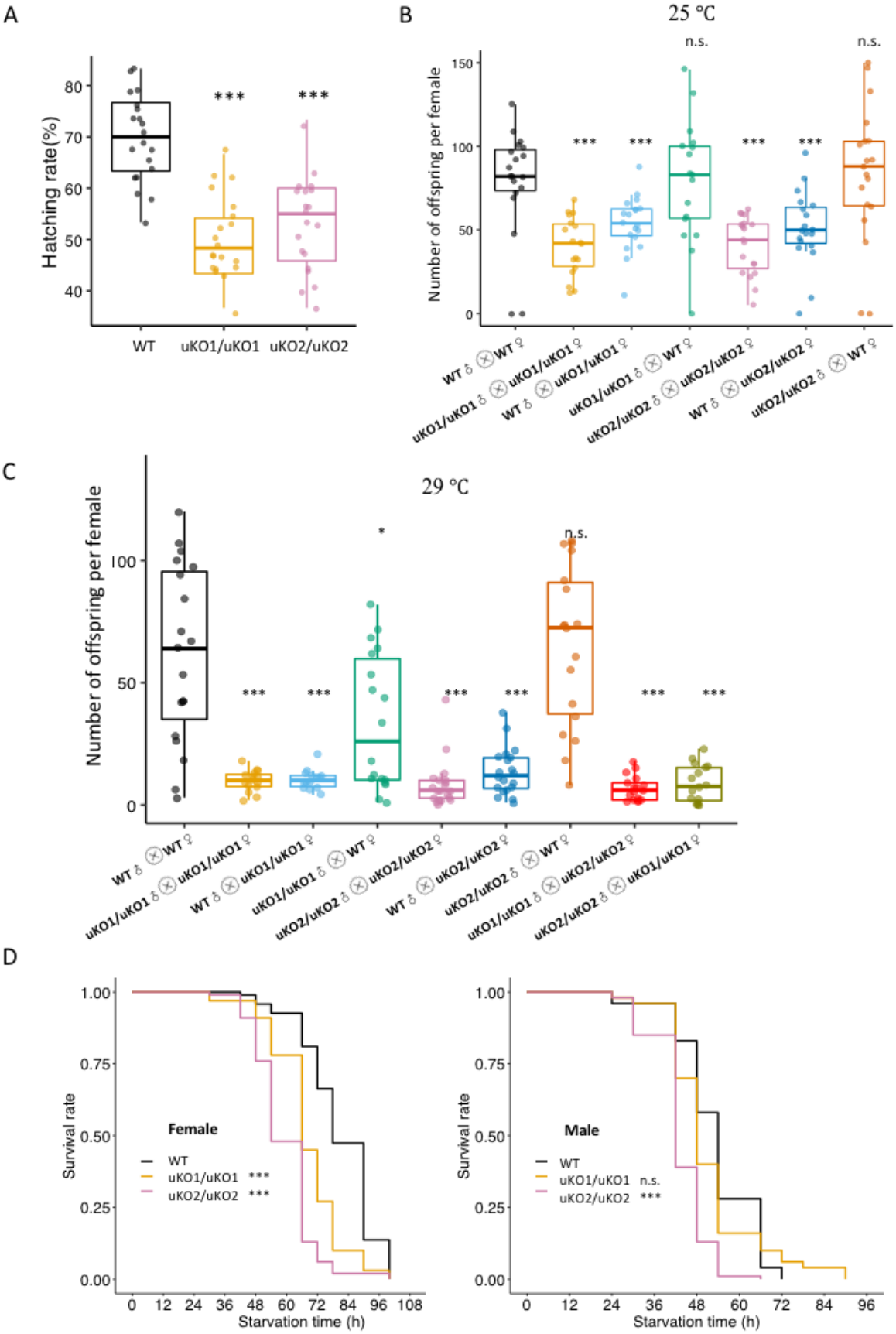


Figure 5. **The strong buffering uORF of *bcd* and it knockout.** (A) Multiple sequence alignment of the *bcd* uORF and partial CDS in *D. melanogaster* and 20 other *Drosophila* species. The uORF and CDS are boxed in green and purple, respectively. The start codons of the uORF and CDS are boxed in red. (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

1240 technology. The uORF is boxed in dark green, and the red ATG represents the start codon of the uORF
1241 in the *D. melanogaster* genome.
1242

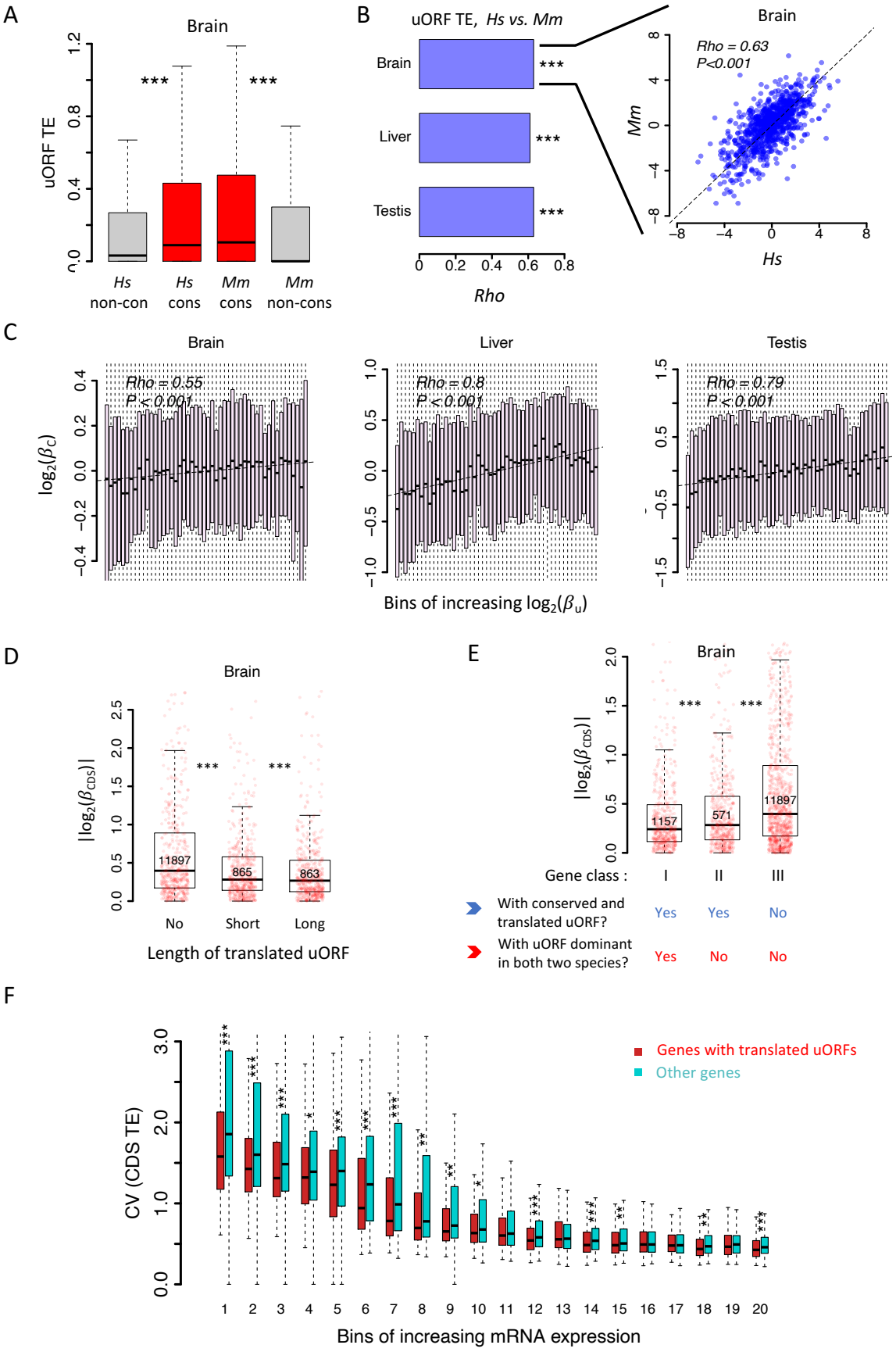

A
uORF
CDS
bcd mRNA
Wild-type: ...CGCAAAAACGCAAAATGTGGGCGAAATAAGTTCGCGAGCGTCTC...
uKO1/uKO1: ...CGCAAA - - T - -AA - - - GTT - - CGAAATAAGTTCGCGAGCGTCTC...
uKO2/uKO2: ...CGCAAAAACGCAAAATAAGTGGGCGAAATAAGTTCGCGAGCGTCTC...
B
Ribosome fractionation
Absorbance (a.u.)
monosome
60S
40S
Polysomes
Monosome fraction
Polysome fraction
C
P-to-M ratio
n.s.
WT
uKO1/uKO1
uKO2/uKO2
25°C
29°C
D
Dual-luciferase reporter
Firefly
Renilla
WT
uKO1
uKO2
ATG uORF
GTT
ATAAG
CDS
Relative luciferase activity
E
25°C
29°C
Number of intersecting DEGs
0–2h embryo
2–6h embryo
6–12h embryo
12–24h embryo
DEG number
F
q-value < 0.05
q-value > 0.05
cell morphogenesis involved in differentiation
embryonic pattern specification
neuron projection development
blastoderm segmentation
mitochondrial translation
regionalization
axonogenesis
segmentation
pattern specification process
translation
carboxylic acid metabolic process
metamorphosis
small molecule biosynthetic process
instar larval or pupal morphogenesis
circulatory system development
epithelial tube morphogenesis
imaginal disc-derived appendage morphogenesis
post-embryonic animal organ development
striated muscle cell development
imaginal disc morphogenesis
cytoplasmic translation
behavior
negative regulation of transcription by RNA polymerase
central nervous system development
head development
axon development
taxis
cell projection morphogenesis
chemotaxis
drug metabolic process
oxidation-reduction process
cellular respiration
mitochondrial gene expression
ATP metabolic process
purine ribonucleoside triphosphate metabolic process
electron transport chain
oxidative phosphorylation
generation of precursor metabolites and energy
0-2h
2-6h
6-12h
12-24h
Embryo

1245 **Figure 6. Knocking out the *bcd* uORF increases CDS translation and perturbs the transcriptome**
1246 **during *D. melanogaster* embryogenesis.** (A) Dual-luciferase assay for *bcd* WT uORF and mutated
1247 uORF. The reporter structures of the WT and uORF mutants are illustrated on the left. The uORF mutant
1248 sequence was the same as that in the fly mutant created with CRISPR-Cas9 technology. The relative
1249 activity of *Renilla* luciferase was normalized to that of firefly luciferase. Error bars represent the S.E.
1250 of six biological replicates. Asterisks indicate statistical significance (***, $P < 0.001$). (B) Two
1251 ribosome fractions (monosome and polysome) of 0-2 h embryos were separated in a sucrose density
1252 gradient. Relative RNA abundance in the monosome and polysome fractions was quantified by real-
1253 time quantitative PCR. (C) P-to-M ratio of *bcd* mRNA (*bcd* mRNA abundance in polysome fraction/*bcd*
1254 mRNA abundance in monosome fraction) at 25°C (left) and 29°C (right). The P-to-M ratios of mutants
1255 were normalized to WT controls at 25°C and at 29°C, respectively. Error bars represent the S.E. of six
1256 biological replicates. Asterisks indicate statistical significance (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$;
1257 n.s., $P > 0.05$). (D) The number of DEGs in each stage and their intersection with each other at 25°C
1258 (top) and 29°C (bottom). (E) Gene ontology analysis of DEGs at 29°C in each stage. The biological
1259 process (BP) terms with q-values < 0.05 in each stage are indicated in red and others are indicated in
1260 white.
1261


A
Hatching rate(%)
80
70
60
50
40
***
***
WT
uKO1/uKO1
uKO2/uKO2
B
25 °C
Number of offspring per female
150
100
50
0
n.s.
n.s.
***
***
***
***
WT ♂ ⊗ WT ♀
uKO1/uKO1 ♂ ⊗ uKO1/uKO1 ♀
WT ♂ ⊗ uKO1/uKO1 ♀
uKO1/uKO1 ♂ ⊗ WT ♀
uKO2/uKO2 ♂ ⊗ uKO2/uKO2 ♀
WT ♂ ⊗ uKO2/uKO2 ♀
uKO2/uKO2 ♂ ⊗ WT ♀
C
29 °C
Number of offspring per female
100
50
0
*
n.s.
***
***
***
***
***
***
WT ♂ ⊗ WT ♀
uKO1/uKO1 ♂ ⊗ uKO1/uKO1 ♀
WT ♂ ⊗ uKO1/uKO1 ♀
uKO1/uKO1 ♂ ⊗ WT ♀
uKO2/uKO2 ♂ ⊗ uKO2/uKO2 ♀
WT ♂ ⊗ uKO2/uKO2 ♀
uKO2/uKO2 ♂ ⊗ WT ♀
uKO1/uKO1 ♂ ⊗ uKO2/uKO2 ♀
uKO2/uKO2 ♂ ⊗ uKO1/uKO1 ♀
D
Survival rate
1.00
0.75
0.50
0.25
0.00
Female
WT
uKO1/uKO1 ***
uKO2/uKO2 ***
0 12 24 36 48 60 72 84 96 108
Starvation time (h)
Survival rate
1.00
0.75
0.50
0.25
0.00
Male
WT
uKO1/uKO1 n.s.
uKO2/uKO2 ***
0 12 24 36 48 60 72 84 96
Starvation time (h)

1262
1263

1264 **Figure 7. Knockout of the *bcd* uORF reduces offspring number and starvation resistance**. (A)
1265 Comparison of the hatching rates (%) of mutant and WT offspring (n=20, Wilcoxon rank sum test; ***,
1266 $P < 0.001$). (B) The offspring number per maternal parent in different crosses over 10 days at 25°C.
1267 Asterisks indicate significant differences between various crosses and crosses of WT females with WT
1268 males (n=20, Wilcoxon rank sum test; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; n.s., $P > 0.05$). The
1269 different crosses were denoted as the x-axis labels. (C) The offspring number per maternal parent in
1270 different crosses over 10 days at 29°C. (D) Survival curves of WT and mutant adult flies of females
1271 (left) and males (right) under starvation conditions. The black line represents the WT, the red line
1272 represents the uKO1/uKO1 mutant, and the blue line represents the uKO2/uKO2 mutant. Asterisks
1273 indicate significant differences compared to the WT. (n=200, log-rank test; ***, $P < 0.001$; n.s., $P >$
1274 0.05).
1275


A
Brain
uORF TE
***
***
Hs non-con
Hs cons
Mm cons
Mm non-cons
B
uORF TE, Hs vs. Mm
Brain
Liver
Testis
Rho
Brain
Rho = 0.63
P<0.001
Mm
Hs
C
Brain
Liver
Testis
Rho = 0.55
Rho = 0.8
Rho = 0.79
P < 0.001
log2(βC)
Bins of increasing log2(βu)
D
Brain
|log2(βCDS)|
11897
865
863
No
Short
Long
Length of translated uORF
E
Brain
11897
1157
571
Gene class : I II III
With conserved and translated uORF? Yes Yes No
With uORF dominant in both two species? Yes No No
F
Genes with translated uORFs
Other genes
CV (CDS TE)
Bins of increasing mRNA expression

1278 **Figure 8. uORFs function as translational buffers in primates.** (A) Boxplots showing the TEs of
1279 conserved and species-specific uORFs between *Hs* (*H. sapiens*) and *Mm* (*M. mulatta*). Data for the
1280 brain is shown as an example. Wilcoxon rank sum tests. ***, $P < 0.001$. (B) Spearman's correlation
1281 coefficient (*Rho*) of uORFs' TE between humans and macaques. The *Rho* values in the brain, liver, and
1282 testis were shown as bar plots. ***, $P < 0.001$. Data for the brain is shown as an example in the right
1283 panel. (C) Correlation between interspecific uORF TE changes ($\log_2\beta_u$) and corresponding CDS TE
1284 changes ($\log_2\beta_C$) in three tissues. The x-axis was divided into 50 equal bins with increasing β_u. (D)
1285 Genes expressed in brains were classified into three classes according to the total length of translated
1286 uORFs. Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P*
1287 values were calculated using Wilcoxon rank sum tests between the neighboring groups. ***, $P < 0.001$.
1288 (E) Genes expressed in brains (mRNA RPKM > 0.1 in both species) were classified into three classes
1289 according to whether a gene had a conserved and dominantly translated uORF (TE > 0.1) in both species
1290 or not. Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P* values
1291 were calculated using Wilcoxon rank sum tests between the neighboring groups. ***, $P < 0.001$. (F)
1292 Boxplot showing the coefficients of variation (CVs) of CDS TE among the 69 lymphoblastoid cell lines
1293 (LCLs). Expressed genes (mean mRNA RPKM > 0.1) were divided into 20 bins with increased mRNA
1294 expression levels. In each bin, the genes were divided into two fractions according to whether the gene
1295 had a translated uORF or not. Wilcoxon rank sum tests. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.
1296