Trans-splicing enhances translational efficiency in *C. elegans*

Yu-Fei Yang,[1,2,3,6] Xiaoqing Zhang,[1,2,3,6] Xuehua Ma,[2,4,6] Taolan Zhao,[1,2,6] Qiushi Sun,[1,2,5] Qing Huan,[1,2] Shaohuan Wu,[1,2,3] Zhuo Du,[4] and Wenfeng Qian[1,2,3]

[1]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; [2]Key Laboratory of Genetic Network Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; [3]University of Chinese Academy of Sciences, Beijing 100049, China; [4]State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; [5]Beijing Key Laboratory of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Translational efficiency is subject to extensive regulation. However, the factors influencing such regulation are poorly understood. In *Caenorhabditis elegans*, 62% of genes are *trans*–spliced to a specific spliced leader (SL1), which replaces part of the native 5′ untranslated region (5′ UTR). Given the pivotal role the 5′ UTR plays in the regulation of translational efficiency, we hypothesized that SL1 *trans*–splicing functions to regulate translational efficiency. With genome-wide analysis on Ribo-seq data, polysome profiling experiments, and CRISPR-Cas9–based genetic manipulation of *trans*–splicing sites, we found four lines of evidence in support of this hypothesis. First, SL1 *trans*–spliced genes have higher translational efficiencies than non-*trans*–spliced genes. Second, SL1 *trans*–spliced genes have higher translational efficiencies than non-*trans*–spliced orthologous genes in other nematode species. Third, an SL1 *trans*–spliced isoform has higher translational efficiency than the non-*trans*–spliced isoform of the same gene. Fourth, deletion of *trans*-splicing sites of endogenous genes leads to reduced translational efficiency. Importantly, we demonstrated that SL1 *trans*–splicing plays a key role in enhancing translational efficiencies of essential genes. We further discovered that SL1 *trans*–splicing likely enhances translational efficiency by shortening the native 5′ UTRs, hence reducing the presence of upstream start codons (uAUG) and weakening mRNA secondary structures. Taken together, our study elucidates the global function of *trans*–splicing in enhancing translational efficiency in nematodes, paving the way for further understanding the genomic mechanisms of translational regulation.

[Supplemental material is available for this article.]

Although proteins are the major macromolecules performing cellular activities, their relative concentrations are largely unknown (Tyers and Mann 2003; Altelaar et al. 2013). Instead, mRNA concentrations are often used as a proxy for protein concentrations in genome-wide studies. This assumes negligible variation of translational efficiencies among genes. With quantitative proteomic data becoming more widely available (Bantscheff et al. 2012; Bensimon et al. 2012), it was surprisingly found that protein and mRNA concentrations are only moderately correlated. For example, only 32% of the variance in protein concentrations in the budding yeast *Saccharomyces cerevisiae* could be explained by mRNA concentration (Ghaemmaghami et al. 2003). Similar phenomena were observed in mice (Schwanhausser et al. 2011) and humans (Wilhelm et al. 2014), in which the proportion of explained variance varied from 0.10 to 0.41 across different tissues. In prokaryotes, where operons are pervasive in the genome, the disparity between mRNA and protein concentrations can become extreme. This is because genes in an operon are transcribed together, yet different protein concentrations may be required to maintain the stoichiometric relationship in protein complexes. Indeed, it was discovered that in *Escherichia coli*, the protein concentration variation in an operon cannot be explained by mRNA concentration (Quax et al. 2013; Li et al. 2014), implying regulation at the level of translation. Furthermore, mRNA concentration differences between haploid and diploid yeast are not an accurate predictor of the differences at the protein level (de Godoy et al. 2008), suggesting the presence of post-transcriptional regulation. Finally, the expression divergence among species is usually smaller at the protein level than that at the mRNA level (Khan et al. 2013), which also suggests the existence of extensive translational regulation (Artieri and Fraser 2014b; McManus et al. 2014).

The process of translation is comprised of three steps: initiation, elongation, and termination (Scheper et al. 2007). Initiation is generally considered the rate-determining step in endogenous genes (Andersson and Kurland 1990; Bulmer 1991; Plotkin and Kudla 2011; Shah et al. 2013; Chu et al. 2014). The 5′ UTR and the 5′ terminus of coding sequences have both been reported to be the main targets for translational regulation (Hall et al. 1982; de Smit and van Duin 1990; Babendure et al. 2006; Kudla et al. 2009; Dvir et al. 2013). Consistently, sequence motifs involved in translational regulation have been identified in the 5′ UTR. Examples include the Kozak sequence (Kozak 1987; Dvir et al. 2013)—the consensus sequence around the start codon

[6]These authors contributed equally to this work.
Corresponding authors: wfqian@genetics.ac.cn, zdu@genetics.ac.cn
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.202150.115.

AUG—and upstream open reading frame (uORF) (Ingolia et al. 2009; Dvir et al. 2013).

Trans-splicing edits 5′ UTR sequences in nematodes, trypanosomes, dinoflagellates, flatworms, and hydra, among many other species (Blumenthal 2004; Lasda and Blumenthal 2011), and thus may be involved in the regulation of translational efficiency. In *Caenorhabditis elegans*, the majority of genes are spliced leader (SL) *trans*-spliced (Blumenthal 2005; Allen et al. 2011). In this process, an SL RNA trims the 5′ UTR of pre-mRNA and then attaches a short (~22 nucleotide [nt]) sequence to the 5′ terminus (Krause and Hirsh 1987; Hastings 2005). After SL *trans*-splicing, ~49% of transcripts retain less than 10 nt of the 5′ UTR sequence of the pre-mRNAs (Lall et al. 2004). SL *trans*-splicing can be classified into two types based on the sequence of the spliced leader: SL1 and SL2, affecting 62% and 12% of genes, respectively (Allen et al. 2011). SL2 *trans*-splicing is related to eukaryotic operons. In *C. elegans*, more than 17% of genes are localized in operons, where they are transcribed into a single nonfunctional polycistronic pre-mRNA (Blumenthal et al. 2002; Blumenthal and Gleason 2003; Blumenthal 2004; Allen et al. 2011; Saito et al. 2013). SL2 *trans*-splicing is used by downstream genes in operons to generate functional monocistronic mRNAs from the pre-mRNA (Spieth et al. 1993; Zorio et al. 1994; Blumenthal et al. 2002). In contrast, SL1 *trans*-splicing is used by the first genes in operons and genes that are not in operons. Although SL1 *trans*-splicing is more prevalent, the function of it is not yet fully understood (Blumenthal 2005). Since SL1 *trans*-splicing modifies the 5′ UTRs of genes, we hypothesized that it functions to regulate translational efficiency (number of proteins made per mRNA per unit time).

Results

SL1 *trans*–spliced genes exhibit higher translational efficiencies than non-*trans*–spliced genes

To test if SL1 *trans*-splicing regulates translational efficiency, we first compared the translational efficiencies of SL1 *trans*-spliced genes and non-*trans*-spliced genes. We calculated ribosome density from previously published ribosome profiling (Ribo-seq) data and the corresponding mRNA-seq data in nematodes (Stadler and Fire 2013) and used this as a measure of translational efficiency (Ingolia et al. 2009). We classified a gene as SL1 *trans*-spliced if ≥90% of informative mRNA-seq reads supported SL1 *trans*-splicing, and a gene as non-*trans*-spliced if no reads supported *trans*-splicing (Allen et al. 2011).

We found that SL1 *trans*-spliced genes had significantly higher translational efficiencies than non-*trans*-spliced genes in *C. elegans* ($P = 1 \times 10^{-214}$, Mann-Whitney *U* test) (Fig. 1A). In fact, the median translational efficiency among SL1 *trans*-spliced genes was approximately twice as much as that of non-*trans*-spliced genes (Fig. 1A). We further identified SL1 *trans*-spliced and non-*trans*-spliced genes in three related nematode species (*C. briggsae*, *C. remanei*, and *C. brenneri*) that diverged from *C. elegans* tens of millions of years ago (Stein et al. 2003; Cho et al. 2004; Hillier et al. 2007; Cutter 2008). We repeated the analysis in these species and observed similar phenomena (Fig. 1B–D). This indicates that the enhanced translational efficiencies in SL1 *trans*-spliced genes are evolutionarily conserved. We further demonstrated that the enhanced translational efficiency of SL1 *trans*-spliced gene is not condition-specific (Supplemental Fig. S1) by comparing translational efficiencies in either L1 diapause or developing worms (Stadler and Fire 2013).

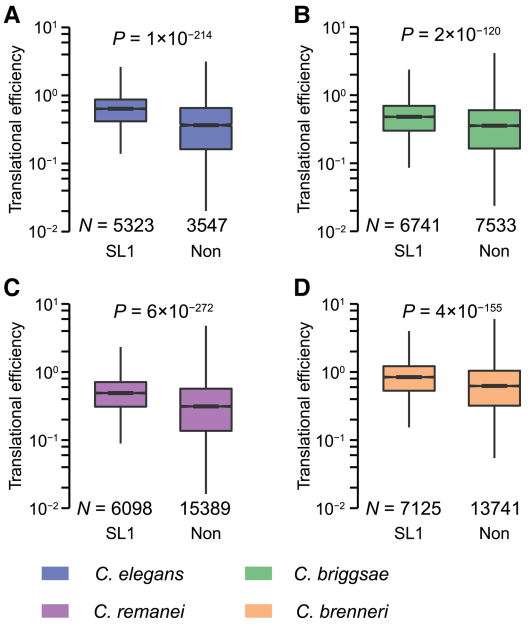

Figure 1. SL1 *trans*-spliced genes exhibit higher translational efficiencies in nematodes. Comparisons were made in (*A*) *C. elegans*, (*B*) *C. briggsae*, (*C*) *C. remanei*, and (*D*) *C. brenneri*. *P*-values were given by the Mann-Whitney *U* test.

It is worth noting that the observed higher translational efficiencies in SL1 *trans*-spliced genes may not be the direct consequence of SL1 *trans*-splicing. For example, the initiation codon AUG may be in a better context in SL1 *trans*-spliced genes, which may precipitate higher translational efficiency (Kozak 1987; Dvir et al. 2013). To examine this possibility, we analyzed only those genes with a perfect Kozak consensus sequence (AAAAUG) (Supplemental Fig. S2A) and again found an approximately two-fold translational efficiency difference between SL1 *trans*-spliced and non-*trans*-spliced genes ($P = 1 \times 10^{-41}$, Mann-Whitney *U* test) (Supplemental Fig. S2B), indicating that the Kozak sequence was not a confounding factor. Furthermore, we calculated an "initiation score" to quantify the optimality of the initiation context around AUG (similarity of the sequence around AUG to the Kozak consensus sequence) in *C. elegans* (see Supplemental Methods). As expected, the initiation score was positively correlated with translational efficiency (Supplemental Fig. S2C). Nevertheless, after controlling for the initiation score, we still observed higher translational efficiencies in SL1 *trans*-spliced genes ($P = 6 \times 10^{-10}$, linear regression model: translational efficiency ~ *trans*-splicing + initiation score) (Supplemental Fig. S2C).

In this study, we used ribosome density as a measure of translational efficiency, assuming similar translational elongation rates among genes. Although it has been reported that synonymous codons of the same amino acid are translated at similar speed (Li et al. 2012; Qian et al. 2012; Artieri and Fraser 2014a; Pop et al. 2014), it has also been argued elsewhere that the elongation rate varied among codons in nematodes, partially due to the different decoding rates between wobble pairing and Watson-Crick pairing at the third position of a codon (Stadler and Fire 2011). Nevertheless, after controlling for the codon adaptation index (CAI), which reflects the synonymous codon usage bias of a gene (Sharp and Li 1987), we still observed significantly higher translational efficiencies among SL1 *trans*-spliced genes ($P = 1 \times 10^{-7}$,

linear regression model: translational efficiency ~ *trans*-splicing + CAI) (Supplemental Fig. S3).

SL1 *trans*–spliced genes exhibit higher translational efficiencies than their non-*trans*–spliced one-to-one orthologs in other nematode species

We further predicted that SL1 *trans*-spliced genes would exhibit higher translational efficiencies than their non-*trans*-spliced one-to-one orthologs in other species. To test this prediction, we first retrieved four-way one-to-one orthologous genes among *C. elegans*, *C. briggsae*, *C. remanei*, and *C. brenneri* (Harris et al. 2014). To avoid false discovery of orthologous genes with *trans*-splicing turnover (see Supplemental Methods; Supplemental Fig. S4), we focused on the one-to-one orthologous genes with SL1 *trans*-splicing evidence in *C. briggsae*, *C. remanei*, or *C. brenneri*, but without *trans*-splicing evidence in *C. elegans*. We found that SL1 *trans*-spliced one-to-one orthologous genes indeed had higher translational efficiencies ($P = 3 \times 10^{-3}$, 0.03, and 1×10^{-4}, respectively, paired Mann-Whitney *U* test) (Fig. 2A–C, left panels). As a control, one-to-one orthologous genes with evidence for SL1 *trans*-splicing in both species did not have a significant difference in translational efficiency ($P = 0.1$, 0.9, and 0.1, respectively, paired Mann-Whitney *U* test) (Fig. 2A–C, right panels).

SL1 *trans*–spliced isoforms exhibit higher translational efficiencies than the non-*trans*–spliced isoforms of the same genes

It has been reported that the *trans*-splicing frequencies of some genes are not 100%, i.e., only a proportion of transcripts are *trans*-spliced (Allen et al. 2011). For these partially *trans*-spliced genes, we predicted that the SL1 *trans*-spliced isoform should be translated more efficiently than the non-*trans*-spliced isoform. To test this prediction, we identified genes with partial SL1 *trans*-splicing in *C. elegans* and calculated the SL1 *trans*-splicing frequency of each gene. Consistent with our prediction, we found that a gene's SL1 *trans*-splicing frequency was positively correlated to its translational efficiency ($\rho = 0.18$, $P = 2 \times 10^{-51}$, Spearman's correlation) (Supplemental Fig. S5).

To further test this hypothesis, we cultured wild type *C. elegans* (N2) in liquid media and performed sucrose gradient centrifugation on the lysate to separate mRNAs into fractions based on the number of ribosomes bound to them (Fig. 3A). Fractions collected later in the experiment (with higher rank) contained more ribosomes per mRNA. Because *trans*-splicing did not change the length of coding sequence, the isoform with more ribosomes had a higher translational efficiency. Thus, for a partially SL1 *trans*-spliced gene, our hypothesis predicted an increasing proportion of the SL1 *trans*-spliced isoform in higher rank fractions. We first randomly chose 12 genes with partial SL1 *trans*-splicing and designed DNA primers (Supplemental Table S1) to specifically quantify isoforms with or without SL1 *trans*-splicing in each fraction with quantitative polymerase chain reactions (qPCR) (Fig. 3B). Consistent with our prediction, in 10 out of 12 genes, we observed reproducible significant positive correlations between the SL1/non-SL ratio and the rank of the fractions (Fig. 3C; Supplemental Fig. S6). In contrast, we observed no significant negative correlations in these 12 genes, validating the function of SL1 *trans*-splicing in promoting translational efficiency ($P = 0.002$, two-tailed binomial test).

Knocking out SL1 *trans*–splicing sites reduces translational efficiency

Given that SL1 *trans*-spliced isoforms exhibited higher translational efficiencies than non-*trans*-spliced isoforms (Fig. 3), we further predicted that knocking out the SL1 *trans*-splicing site of a gene should reduce its translational efficiency and thus its protein concentration. To quantify protein concentration, we used three protein::GFP fusion reporters (Supplemental Table S2) in which a gene encoding green fluorescent protein (*GFP*) had been fused to the 3′ end of an SL1 *trans*-spliced endogenous gene (Paix et al. 2014). We next attempted to knock out sequences that are important for SL1 *trans*-splicing frequency. To this end, we retrieved sequences around all the annotated SL1 *trans*-splicing sites (Allen et al. 2011) and identified their consensus sequence TTNCAG (Fig. 4A) by WebLogo v2.8 (Schneider and Stephens 1990; Crooks et al. 2004), which is in accordance with previous findings (Conrad et al. 1993). We confirmed that the SL1 *trans*-splicing sites with this consensus sequence exhibited significantly higher *trans*-splicing frequencies ($P = 5 \times 10^{-192}$, Mann-Whitney *U* test) (Fig. 4B). We then knocked out the consensus sequence of the *trans*-splicing sites (TTTCAG) (Fig. 4C) in these GFP fusion strains with the CRISPR-Cas9 approach. After confirming that the SL1 *trans*-splicing frequency was indeed reduced in SL1 mutants (see Methods; Supplemental Fig. S7A–D), we compared the GFP intensity between the wild type and SL1 mutants at a comparable developmental stage. Consistent with our prediction, GFP intensities were reduced to 50%–75% in all three SL1 mutants ($P = 2 \times 10^{-21}$, 1×10^{-11}, and 4×10^{-7}, respectively, Mann-Whitney *U* test) (Fig. 4D–F), while the mRNA levels remained largely unchanged (Supplemental Fig. S7E–G), confirming our prediction that decreased *trans*-splicing frequency leads to reduced translational efficiency.

SL1 *trans*-splicing-mediated translational enhancement is pivotal for fitness. For example, *deps-1* encodes a P-granule-associated protein and is required for embryonic viability and fertility (Spike et al. 2008). Although the coding sequence of *deps-1* remained unchanged, the deletion of the SL1 *trans*-splicing site (Fig. 4C,F; Supplemental Fig. S7D,G) was sufficient

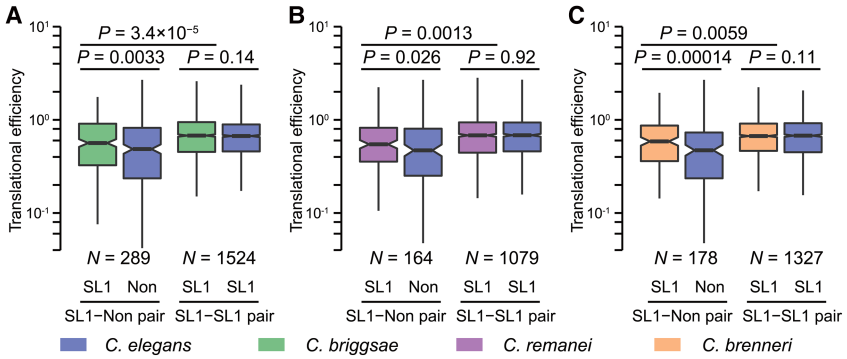

Figure 2. SL1 *trans*-spliced genes exhibit higher translational efficiencies than their non-*trans*-spliced one-to-one orthologs. Comparisons were made between *C. elegans* and (*A*) *C. briggsae*, (*B*) *C. remanei*, and (*C*) *C. brenneri*. *P*-values were given by the Mann-Whitney *U* test.

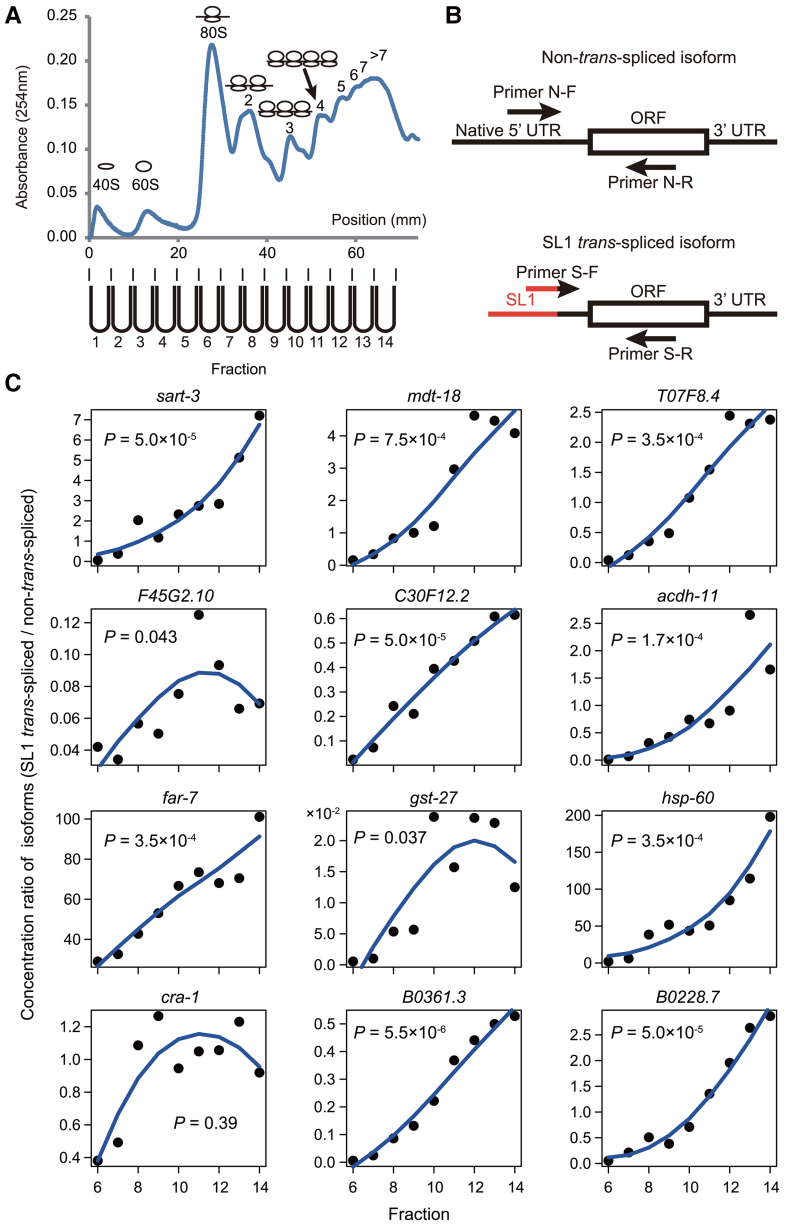

Figure 3. SL1 *trans*-spliced isoforms exhibit higher translational efficiencies than the non-*trans*-spliced isoforms of the same genes. (*A*) mRNAs with different numbers of ribosomes bound to them were fractionated by sucrose gradient centrifugation. (*B*) Isoform-specific qPCR primers were designed to quantify the concentrations of two isoforms. (*C*) The concentration ratio of two isoforms (SL1 *trans*-spliced/non-*trans*-spliced) was quantified in each fraction of fractions 6–14. The locally weighted scatterplot smoothed (LOWESS) lines are shown in blue. *P*-values were given by Spearman's rank correlation.

to induce partially penetrant embryonic lethality, reduced fertility, and decreased lifetime fecundity ($P = 0.01$, 0.03, and 5×10^{-7}, respectively, *t*-test) (Fig. 4G–I), phenocopying the loss-of-function mutant (Spike et al. 2008) and highlighting the indispensable role of *trans*-splicing in worm development.

Translational efficiencies of essential genes are enhanced by SL1 *trans*–splicing

Essential genes are defined as having lethal or sterile phenotypes upon gene perturbation or knockdown (Kamath et al. 2003). Interestingly, we found that essential genes also exhibited higher translational efficiencies than nonessential genes in *C. elegans* ($P = 6 \times 10^{-6}$, Mann-Whitney *U* test) (Fig. 5A). To investigate if this difference is attributable to SL1 *trans*-splicing, we first calculated the proportion of essential genes among SL1 *trans*-spliced and non-*trans*-spliced genes, respectively. We found that essential genes were significantly enriched among SL1 *trans*-spliced genes ($P = 1 \times 10^{-82}$, Fisher's exact test) (Fig. 5B). We further observed that essential and nonessential genes had virtually equal translational efficiencies within both SL1 *trans*-spliced and non-*trans*-spliced groups ($P = 0.03$ and 0.18, respectively, Mann-Whitney *U* test) (Fig. 5C). More importantly, nonessential genes with SL1 *trans*-splicing even exhibited higher translational efficiencies than essential genes without *trans*-splicing ($P = 2 \times 10^{-7}$, Mann-Whitney *U* test) (Fig. 5C). Taken together, the preference of SL1 *trans*-splicing among essential genes largely explains the difference in translational efficiency between essential and nonessential genes.

It is worth noting that the preference of essential genes to be SL1 *trans*-spliced might be a byproduct of the enrichment of essential genes in operons (Blumenthal and Gleason 2003; Zaslaver et al. 2011). To examine this possibility, we further divided genes into operon genes and nonoperon genes and found that essential genes were more likely to be SL1 *trans*-spliced among both operon and nonoperon genes (Fig. 5B), indicating that the enrichment of essential genes in SL1 *trans*-spliced genes is independent of operon status.

SL1 *trans*–splicing trims the 5′ UTR to enhance translational efficiency

How is translational efficiency increased through SL1 *trans*-splicing? It has been reported that a long 5′ UTR reduces translational efficiency (Chappell et al. 2006; Staley et al. 2012; Paek et al. 2015). Therefore, we hypothesized that *trans*-splicing increases translational efficiency by reducing the 5′ UTR length. To test the hypothesis, we first confirmed that translational efficiency was indeed negatively correlated with 5′ UTR length in *C. elegans*, among both SL1 *trans*-spliced (Fig. 6A) and non-*trans*-spliced genes (Fig. 6B). Importantly, we found that before SL1 *trans*-splicing, the 5′ UTRs of SL1 *trans*-spliced genes (pre-SL1 transcripts) were longer than those of non-*trans*-spliced genes ($P = 5 \times 10^{-20}$, Mann-Whitney *U* test) (Fig. 6C). SL1 *trans*-splicing significantly shortens 5′ UTRs ($P = 1 \times 10^{-62}$, Mann-Whitney *U* test) (Fig. 6C), making them even

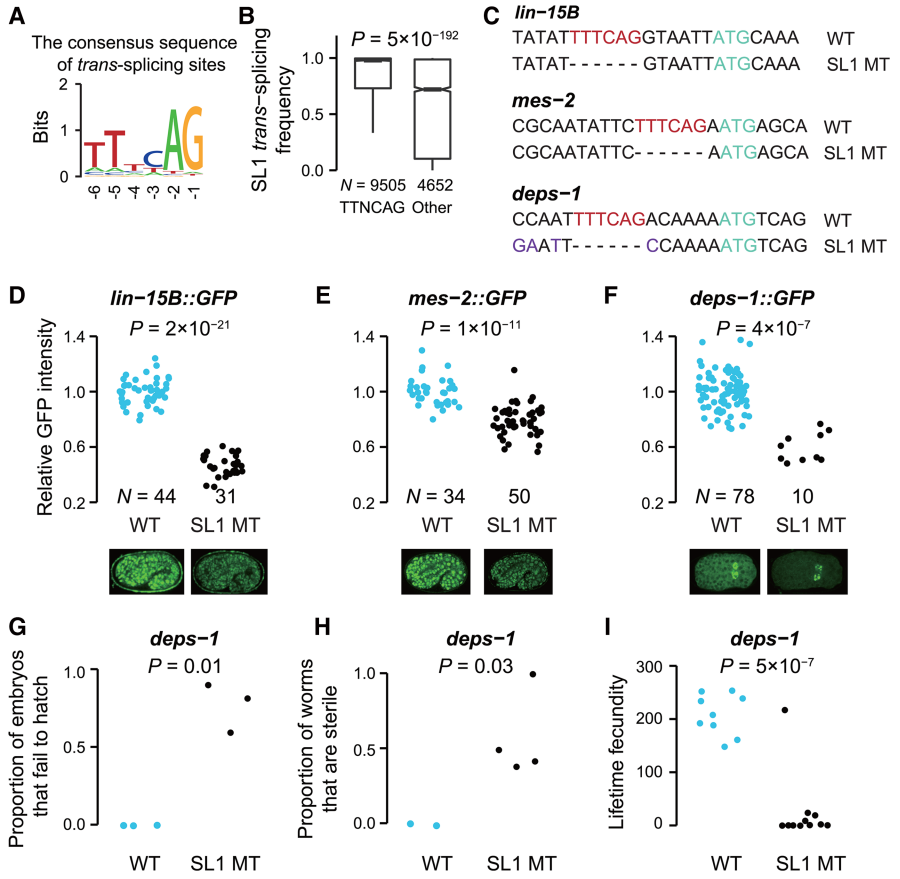

Figure 4. Knocking out the consensus sequence of *trans*-splicing sites leads to reduced translational efficiency. (*A*) The consensus sequence of SL1 *trans*-splicing sites in *C. elegans* (TTNCAG). (*B*) The consensus sequence significantly increases the efficiency of SL1 *trans*-splicing ($P = 5 \times 10^{-192}$, Mann-Whitney *U* test). (*C*) The sequences of wild type (WT) and SL1 *trans*-splicing mutants (SL1 MT) around the *trans*-splicing sites are shown for *lin-15B*, *mes-2*, and *deps-1*. The consensus sequences of *trans*-splicing sites are marked in red and the start codons are marked in cyan. Silent mutations were made (in purple) to prevent Cas9 from recutting. (*D–F*) The protein abundances of LIN-15B (*D*), MES-2 (*E*), and DEPS-1 (*F*) were quantified by the fluorescence intensity of GFP. *P*-values were given by the Mann-Whitney *U* test. Because knocking out the *trans*-splicing site of *deps-1* frequently resulted in embryonic lethality and a complete loss of GFP signal, only embryos with GFP signal were used for quantification. An image example of each strain is shown. (*G–I*) Knocking out the *trans*-splicing site of *deps-1* induced partially penetrant embryonic lethality (*G*), reduced fertility (*H*), and decreased lifetime fecundity (*I*). Each dot represents the result of an independent experimental replication. *P*-values were given by the *t*-test.

shorter than those of non-*trans*-spliced genes ($P = 2 \times 10^{-10}$, Mann-Whitney *U* test) (Fig. 6C). This may partly explain the higher translational efficiencies of SL1 *trans*-spliced genes than those of non-*trans*-spliced genes (Fig. 1).

SL1 *trans*–splicing removes uAUGs in 5′ UTRs to enhance translational efficiency

It has been reported that an uAUG may interfere with canonical translational initiation sites and thus reduces translational efficiency (Ingolia et al. 2009; Dvir et al. 2013). Indeed, we found that genes with uAUG had significantly lower translational efficiencies among both SL1 *trans*-spliced genes (fold change of median = 1.8, $P = 4 \times 10^{-62}$, Mann-Whitney *U* test) (Fig. 6D) and non-*trans*-spliced genes (fold change of median = 1.4, $P = 4 \times 10^{-7}$, Mann-Whitney *U* test) (Fig. 6D). In addition, genes with an out-of-frame uAUG had lower translational efficiencies than those with an in-frame uAUG ($P = 7 \times 10^{-5}$, Mann-Whitney *U* test). Furthermore, genes with an in-frame uORF that terminates at an upstream stop codon (USC) were translated less efficiently than those with an in-frame uORF that shares the stop codon (SSC) with the annotated ORF ($P = 3 \times 10^{-6}$, Mann-Whitney *U* test) (Fig. 6E) and were translated with similar efficiencies to genes with an out-of-frame uORF ($P = 0.7$, Mann-Whitney *U* test) (Fig. 6E). This observation suggests that an uAUG may decrease translational efficiency by early release of ribosomes. Given that the SL1 sequence does not contain AUG, SL1 *trans*-splicing can potentially eliminate an uAUG in the 5′ UTR by trimming it. Indeed, we found that SL1 *trans*-splicing increased the proportion of genes without an uAUG ($P = 2 \times 10^{-87}$, Fisher's exact test) (Fig. 6F), making it even higher than that of non-*trans*-spliced genes ($P = 5 \times 10^{-34}$, Fisher's exact test) (Fig. 6F).

SL1 *trans*–splicing attenuates 5′ UTR secondary structure to enhance translational efficiency

Stable secondary structures in the 5′ terminus of an mRNA can slow down ribosome scanning or translocation at the initiation step and thus reduce translational efficiency (Sen et al. 2015). Therefore, we hypothesized that SL1 *trans*-splicing, which replaces part of the native 5′ UTR, might increase translational efficiency by attenuating secondary structures in the 5′ UTR.

We first confirmed that stable secondary structures in the 5′ UTR reduced translational efficiency in *C. elegans*. Specifically, we calculated the minimum free energy (MFE) (Lorenz et al. 2011) for all 22-nt windows in the 5′ UTR for each gene and used the smallest MFE (sMFE) to describe the stability of the most stable secondary structure in a given 5′ UTR. The length of 22 nt was used because it is the shortest 5′ UTR length after SL1 *trans*-splicing. As expected, the sMFE was positively correlated with translational efficiency among both SL1 *trans*-spliced genes [$\rho = 0.18$, $P = 3 \times 10^{-25}$, Spearman's correlation, slope $b = 0.02$, linear regression of $\log_{10}(\text{TE}) \sim \text{sMFE}$] (Fig. 6G) and non-*trans*-spliced genes [$\rho = 0.13$, $P = 3 \times 10^{-5}$, Spearman's correlation, $b = 0.03$, linear regression of $\log_{10}(\text{TE}) \sim \text{sMFE}$] (Fig. 6H).

We found that SL1 *trans*-splicing significantly elevated the sMFE (Fig. 6I). Before SL1 *trans*-splicing, the difference between the sMFE values of pre-SL1 *trans*-spliced transcripts and non-*trans*-spliced transcripts was small ($P = 0.002$, Mann-Whitney *U* test) (Fig. 6I). After SL1 *trans*-splicing, however, the values of the sMFE of SL1 *trans*-spliced transcripts were significantly increased ($P = 1 \times 10^{-22}$, two-tailed Mann-Whitney *U* test) (Fig. 6I), above those of non-*trans*-spliced transcripts ($P = 7 \times 10^{-7}$, Mann-Whitney *U* test) (Fig. 6I).

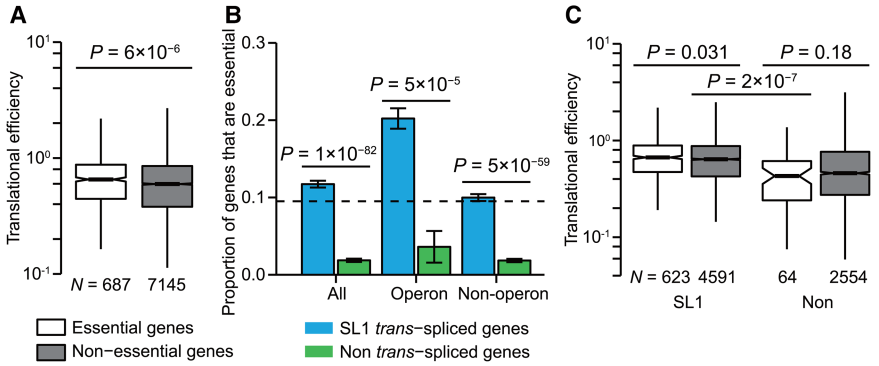

Figure 5. Translational efficiencies of essential genes are enhanced by SL1 *trans*-splicing. (*A*) Essential genes exhibit higher translational efficiencies than nonessential genes (*P*-value was given by the Mann-Whitney *U* test). (*B*) The proportion of genes that are essential is higher in SL1 *trans*-spliced genes (*P*-values were given by Fisher's exact test). This observation holds when genes were separated into two groups: within operons and out of operons. (*C*) After controlling for the type of *trans*-splicing, translational efficiencies are similar between essential and nonessential genes (*P*-values were given by the Mann-Whitney *U* test).

uAUG and mRNA secondary structure largely explain the regulatory effect of 5′ UTR length on translational efficiency

We next asked if the removal of an uAUG and the attenuation of secondary structures in the 5′ UTR can explain all the contributions of 5′ UTR shortening to the enhancement of translational efficiency. To this end, we built various linear regression models and calculated Akaike information criterion (AIC) to examine the relative contributions of different sequence properties in explaining translational efficiency (Table 1). We found that adding 5′ UTR length to the null model significantly reduced the AIC (model 5 vs. model 1). In contrast, after including both the presence/absence of the uAUG and the stability of secondary structures (model 4), adding 5′ UTR length to the model no longer reduced the AIC (model 6), indicating that shortening of the 5′ UTR increased translational efficiency mainly through regulation of the uAUG and 5′ UTR secondary structures in *C. elegans*.

We noted that adding *trans*-splicing type (here defined as SL1 *trans*-splicing or not) to model 4 continued to reduce the AIC (model 8), suggesting additional molecular mechanisms to enhance translational efficiency by SL1 *trans*-splicing. For example, SL1 *trans*-splicing replaces the canonical monomethylguanosine (m^7GpppN, MMG) cap with a trimethylguanosine ($m_3^{2,2,7}$GpppN, TMG) cap, which may also contribute to the regulation of translational efficiency (Maroney et al. 1995). Nevertheless, non-*trans*-spliced genes without an uAUG exhibited higher translational efficiencies than SL1 *trans*-spliced genes that contained an uAUG ($P = 1 \times 10^{-5}$, Mann-Whitney *U* test) (Fig. 6D), suggesting a minor role of the TMG cap compared with the uAUG in the regulation of translation.

In Figure 4, we observed reduced translational efficiency in SL1 mutants. To examine whether SL1 *trans*-splicing increases translational efficiency by removing the inhibitory uAUG or secondary structure from the pre-mRNA, we performed 5′ RACE (rapid amplification of 5′ complementary DNA ends) in each SL1 mutant to obtain the 5′ UTR sequence of pre-mRNA (Supplemental Table S3) and compared it with the SL1 *trans*-spliced transcript. Although the SL1 *trans*-splicing in *lin-15B* did not significantly change the sMFE in the 5′ UTR, it removed three out-of-frame uAUGs and an in-frame uORF that terminates at a USC (Table 2). The difference in GFP intensity between the wild type and the SL1 mutant of *lin-15B* (fold change of median = 2.2) (Fig. 4D) is largely consistent with the average effect of uORFs observed in Figure 6D (1.4- to 1.8-fold). Note that each individual uAUG may have its specific features (such as Kozak context of this uAUG) that make its effect different from the average effect estimated in Figure 6D. SL1 *trans*-splicing also removed an out-of-frame uAUG in *mes-2* (Table 2), largely explaining the difference in GFP intensity between the wild type and the SL1 mutant (fold change of median = 1.3) (Fig. 4E). SL1 *trans*-splicing in this gene also increased the sMFE by 0.6 kcal/mol (Table 2), which may also contribute to the higher translational efficiency in the wild type. In *deps-1*, there is no uORF in pre-mRNA. Nevertheless, SL1 *trans*-splicing attenuated secondary structures by 2 kcal/mol in this gene (Table 2). Based on the linear regression models in Figure 6, G and H, this attenuation on average could lead to an ~1.1-fold increase of translational efficiency, which is smaller than the observed difference in GFP intensity between the wild type and the SL1 mutant of this gene (fold change of median = 1.8) (Fig. 4E). It is likely that other factors such as the TMG cap and the SL1 sequence (Maroney et al. 1995) also contribute to the regulation of translational efficiency in this gene.

Discussion

SL2 *trans*–spliced genes also exhibit higher translational efficiencies

In addition to the major role SL2 *trans*-splicing plays in converting a polycistronic pre-mRNA into multiple monocistronic mRNAs, SL2 *trans*-splicing also attaches a 22-nt sequence to the monocistronic mRNA, which becomes part of the 5′ UTR. We speculated that SL2 might also be involved in the regulation of translational efficiency. To examine this possibility, we compared the translational efficiencies of SL1 *trans*-spliced genes, SL2 *trans*-spliced genes, and non-*trans*-spliced genes (Supplemental Fig. S1). Although less efficiently translated than SL1 *trans*-spliced genes ($P = 1 \times 10^{-11}$, Mann-Whitney *U* test) (Supplemental Fig. S1C), SL2 *trans*-spliced genes exhibited higher translational efficiencies than non-*trans*-spliced genes ($P = 2 \times 10^{-19}$, Mann-Whitney *U* test) (Supplemental Fig. S1C). Similar results were obtained in other nematode species ($P = 4 \times 10^{-5}$, 3×10^{-31}, 2×10^{-5}, in *C. briggsae*, *C. remanei*, and *C. brenneri*, respectively, Mann-Whitney *U* test) (Supplemental Fig. S1F,I,L). These findings suggest the additional role of SL2 *trans*-splicing in translational regulation. The difference in translational efficiency between SL2 *trans*-spliced genes and non-*trans*-spliced genes was highly significant in developing worms (Supplemental Fig. S1). The significance decreased in L1 diapause, likely due to the reduced transcriptional activities of operon genes in growth-arrested states (Zaslaver et al. 2011), which increased the experimental errors in quantifying translational efficiency of SL2 *trans*-spliced genes.

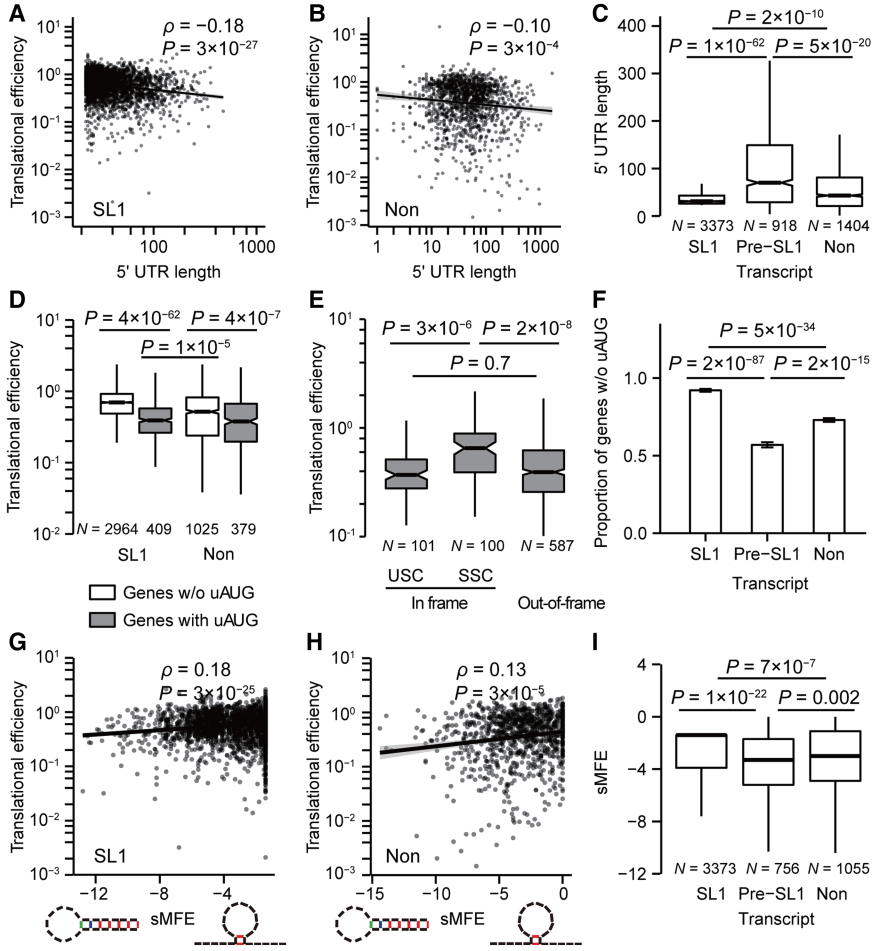

Figure 6. SL1 *trans*-splicing removes the uAUG and attenuates secondary structures in the 5′ UTR to enhance translational efficiency. The 5′ UTR length is negatively correlated with translational efficiency in both (*A*) SL1 *trans*-spliced genes ($N = 3373$) and (*B*) non-*trans*-spliced genes ($N = 1404$). Linear regression lines and their 95% confidence intervals are shown. (*C*) Before *trans*-splicing, the 5′ UTRs of SL1 *trans*-spliced genes (pre-SL1) are longer than those of non-*trans*-spliced genes (Non), whereas after *trans*-splicing, they (SL1) are shorter than those of non-*trans*-spliced genes. *P*-values were given by the Mann-Whitney *U* test. (*D*) Genes without an uAUG exhibit higher translational efficiencies among both SL1 *trans*-spliced and non-*trans*-spliced genes. *P*-values were given by the Mann-Whitney *U* test. (*E*) The translational efficiencies of genes with at least one in-frame uORF that terminates at an upstream stop codon (USC) are significantly lower than those with in-frame uORFs that share the stop codons (SSC) with the annotated ORFs (*P*-value was given by the Mann-Whitney *U* test). In addition, genes with an in-frame uAUG exhibit higher translational efficiencies than those with an out-of-frame uAUG ($P = 7 \times 10^{-5}$, Mann-Whitney *U* test). (*F*) The proportion of genes without an uAUG in non-*trans*-spliced genes is larger than that in pre-SL1 transcripts but is smaller than that in SL1 transcripts (Fisher's exact test). Error bars represent standard errors estimated from a binomial distribution. (*G*,*H*) The smallest MFE (sMFE) was used to estimate the free energy (in units of kcal/mol) of the most stable secondary structure in the 5′ UTR. The sMFE is positively correlated with translational efficiency in both (*G*) SL1 *trans*-spliced genes ($N = 3373$) and (*H*) non-*trans*-spliced genes ($N = 1055$). Schematic secondary structures are shown on the *x*-axis. (*I*) The sMFE in non-*trans*-spliced genes is slightly higher than that in pre-SL1 transcripts but is much smaller than that in SL1 transcripts. *P*-values were given by the Mann-Whitney *U* test.

SL1 *trans*–splicing provides a great in vivo system for studying the molecular mechanisms of translational regulation

Regulation at the translation level has attracted increasing interest over the past few years (de Godoy et al. 2008; Ingolia et al. 2009, 2013; Kudla et al. 2009; Tuller et al. 2010; Stadler and Fire 2011, 2013; Bantscheff et al. 2012; Bensimon et al. 2012; Li et al. 2012, 2014; Qian et al. 2012; Khan et al. 2013; Quax et al. 2013; Chu et al. 2014; McManus et al. 2014; Pop et al. 2014). Although it is widely accepted that the 5′ UTR is the major target for translational regulation, the exact regulatory mechanisms remain elusive. The Segal group addressed this issue by systematically generating point mutations in a 10-nt region in a 5′ UTR and identified a number of sequence properties that regulate translational efficiency (Dvir et al. 2013). However, it remains unclear if the length of the 5′ UTR regulates the translational initiation efficiency. To study the regulatory role of 5′ UTR length, one could generate multiple 5′ UTR sequence variants with various lengths while keeping the other relevant factors constant for the same reporter gene. However, such experiments are labor-intensive, and it is unclear whether the conclusion would be applicable to other genes. Fortunately, *trans*-splicing in nematodes spontaneously "conducted" such experiments at the genomic scale. In this study, we took advantage of this system and identified molecular mechanisms regulating translational efficiency.

In fact, studying the effect of *trans*-splicing on translation regulation has a longstanding history. The first study attempted to address this question with one reporter gene in a cell-free translation system extracted from *Ascaris lumbricoides*. They found that a RNA molecule with both SL1 sequence and a TMG cap translated more efficiently than one with neither (Maroney et al. 1995). However, an opposite conclusion was reached in another in vitro translation system extracted from *A. suum*, which exhibited cap-tail synergism (Lall et al. 2004). The latter group further performed a particle bombardment assay to transiently express reporter RNA in *Ascaris* embryos (Cheng et al. 2007) and reported similar results as Lall et al. (2004). However, this pattern was no longer observed in a more recent paper from the same group (Wallace et al. 2010). The discrepancy among studies may be attributed to the unstable (in vitro translation and transient expression) system and/or the limited number of artificially designed reporter sequences.

SL1 *trans*-splicing leads to a number of consequences for endogenous genes, including but not limited to (1) the switch from an MMG cap to a TMG cap, (2) the addition of the SL1 sequence, (3) the removal of uAUGs, (4) the attenuation of mRNA secondary structures, and (5) the proteins binding to mRNA during *trans*-splicing that may be cotransported to the cytoplasm. Previous studies mainly focused on (1) and (2) for artificially synthesized RNA (Maroney et al. 1995; Lall et al. 2004; Cheng et al. 2007; Wallace et al. 2010). However, the overall effect of

Table 1. Models on sequence properties that regulate translational efficiency (TE)

	Model	AIC
1	Null model	9933
2	TE ~ uAUG[a]	3113
3	TE ~ sMFE[b]	3319
4	TE ~ uAUG + sMFE	3080
5	TE ~ 5′ UTR length	3219
6	TE ~ uAUG + sMFE + 5′ UTR length	3079
7	TE ~ SL type[c]	3269
8	TE ~ uAUG + sMFE + SL type	3008

[a]One for genes with at least one uAUG; 0 for genes without any uAUG.
[b]The free energy of the most stable secondary structure in the 5′ UTR.
[c]One for SL1 *trans*-spliced genes; 0 for non-*trans*-spliced genes.

SL1 *trans*-splicing ([1] to [5]) remains unclear for endogenous genes, especially at the genomics scale. The quantification of this effect is of central importance to understand why SL1 *trans*-splicing is pervasive in nematode genomes. In this study, we provide four lines of evidence supporting that SL1 *trans*-splicing enhances translational efficiencies of endogenous genes.

SL1 *trans*–splicing also contributes to the environmental robustness of translational efficiency in essential genes

Essential genes are more likely to participate in basic biological processes and therefore require more robust translational efficiencies across variable conditions. We compared the fold change of translational efficiency during the exit from L1 diapause (Stadler and Fire 2013) and found that the translational efficiencies of essential genes were more stable during this process ($P = 2 \times 10^{-21}$, *F*-test) (Supplemental Fig. S8A). We speculated that such robustness might be partially attributable to differential SL1 *trans*-splicing between essential and nonessential genes. To test this hypothesis, we compared the translational robustness between SL1 *trans*-spliced and non-*trans*-spliced genes and found that this difference was even larger ($P = 6 \times 10^{-137}$, *F*-test) (Supplemental Fig. S8B), implicating the role that SL1 *trans*-splicing plays in maintaining environmental robustness of translational efficiency. Similar phenomena were also observed in other nematode species (Supplemental Fig. S9A–C). Because essential genes are enriched in SL1 *trans*-spliced genes (Fig. 5B), SL1 *trans*-splicing should contribute to the elevated environmental robustness of translational efficiency among these essential genes. We further compared the translational robustness among four groups of genes based on gene essentiality and *trans*-splicing type in *C. elegans* (Supplemental Fig. S8C) and found that essential genes had stronger robustness among both SL1 *trans*-spliced and non-*trans*-spliced genes ($P = 8 \times 10^{-6}$ and 6×10^{-3}, respectively, *F*-test) (Supplemental Fig. S8C), suggesting that additional mechanisms must exist to regulate the translational robustness of essential genes.

Evolutionary model of SL1 *trans*-splicing

In previous studies, it has been argued that gains of operons are nearly neutral, whereas losses of operons and SL2 *trans*-splicing are extremely deleterious, because the downstream genes in operons have lost promoters (Lawrence 1999; Nimmo and Woollard 2002; Blumenthal 2004). Based on this idea, we previously proposed a quantitative model called "easy come, slow go" to explain the evolution of operons in nematodes (Qian and Zhang 2008). In the present study, we found genomic evidence supporting the hypothesis that SL1 *trans*-splicing promotes translational efficiency by removing the uAUG and attenuating secondary structures in the 5′ UTR (Fig. 6; Table 2). This may also be partly explained by the "easy come, slow go" model. That is, the gain of SL1 *trans*-splicing in a gene is nearly neutral, but once an SL1 *trans*-splicing site is obtained, the accumulation of deleterious mutations (such as uAUGs and stable secondary structures) in the 5′ UTR of the pre-*trans*-spliced transcripts, which are no longer visible to natural selection (Blumenthal 2005), is accelerated. Consistently, the proportion of pre-SL1 transcripts that are without an uAUG was lower than that of the transcripts from non-*trans*-spliced genes (Fig. 6F). In addition, pre-SL1 transcripts contained more stable secondary structures than the transcripts from non-*trans*-spliced genes (Fig. 6I). These uAUGs and stable secondary structures prevent efficient translation and make SL1 *trans*-splicing sites under purifying selection.

However, there are a number of phenomena in the evolution of SL1 *trans*-splicing that cannot be explained by the "easy come, slow go" model. First, the translational efficiencies of SL1 *trans*-spliced genes are higher than non-*trans*-spliced genes (Fig. 1; Supplemental Fig. S1). Second, essential genes tend to be SL1 *trans*-spliced (Fig. 5B). Third, the proportion of SL1 transcripts that are without an uAUG is higher than that of the transcripts from non-*trans*-spliced genes (Fig. 6F). Fourth, SL1 transcripts contain less stable secondary structures than the transcripts from non-*trans*-spliced genes (Fig. 6I). All these observations can be explained by positive selection for incorporation of SL1 *trans*-splicing sites on the genes requiring higher translational efficiencies (such as essential genes). Taken together, our observations imply that SL1 *trans*-splicing contributed to molecular adaptation in nematode evolution.

Methods

Worm strains

Three *C. elegans* transgenic strains JH3203, JH3205, and JH3207 were ordered from the *Caenorhabditis* Genetics Center. In these strains, *GFP* had been fused to the 3′ terminus of endogenous genes *mes-2*, *lin-15B*, and *deps-1*, respectively. These three genes were chosen because (1) they are SL1 *trans*-spliced, and (2) their protein::GFP fusion reporters were available. Liquid culture of N2 worms was performed following Koelle and Herman's protocol.

Table 2. SL1 *trans*-splicing removed uAUGs or weakened mRNA structures in the 5′ UTRs of reporter genes

Gene	Isoform	5′ UTR length	Positions of out-of-frame uAUGs[a]	Positions of in-frame uORFs with USC[b]	sMFE (kcal/mol)
lin-15B	Pre-SL1	346 nt	−43, −46, −143	[−258, −241]	−5.3
	SL1	28 nt	NA	NA	−5.4
mes-2	Pre-SL1	108 nt	−101	NA	−2
	SL1	23 nt	NA	NA	−1.4
deps-1	Pre-SL1	125 nt	NA	NA	−3.4
	SL1	28 nt	NA	NA	−1.4

[a]Positions of base A in out-of-frame uAUGs. The position of base A in the annotated AUG was defined as 0. NA represents the absence of an out-of-frame uAUG in the 5′ UTR.
[b]Positions of the first and the last bases of an in-frame uORF that terminates at an upstream stop codon (USC).

Specifically, worms were grown at 20°C for 4 d in liquid S complete media supplemented with *E. coli* X1666 (Kenyon 1997).

Sucrose gradient centrifugation and fraction collection

The frozen worms were crushed with a mortar and pestle in a liquid nitrogen bath, and the resulting powder was transferred to three volumes of ice-cold polysome lysis buffer (200 mM Tris pH 8.0, 200 mM KCl, 35 mM $MgCl_2$, 1% Triton X-100, 5 mM DTT, 100 μg/mL cycloheximide, and 0.5 mg/mL heparin). Sucrose gradient ultracentrifugation and fraction collection were performed largely as described in a previous study (Ingolia et al. 2009). The detailed protocol is provided in the Supplemental Methods.

Generation and analysis of SL1 *trans*-splicing mutants

We used the CRISPR-Cas9 system to remove the consensus sequence of *trans*-splicing sites of the *GFP* fusion genes in strains JH3203, JH3205, and JH3207, respectively. Two guide sequences were designed for each gene (Supplemental Table S4). To construct the sgRNA expression plasmids, the guide sequences were inserted into the pDD162 plasmid by PCR (Supplemental Table S5) following previous protocols (Paix et al. 2014). Repair template sequences (Supplemental Table S6) were used to remove *trans*-splicing sites (Fig. 4C). Mutations were confirmed by single worm PCR (Supplemental Table S7) and Sanger sequencing. Transcript abundances and *trans*-splicing frequencies of reporter genes were quantified by qPCR (Supplemental Table S8). Embryos were mounted following a well-established protocol with minor modifications (Bao and Murray 2011). Fluorescence micrographs were taken using spinning disk confocal microscopy (PerkinElmer) with Hamamatsu Flash4.0 and 9100-23B cameras. A detailed protocol is described in the Supplemental Methods.

Rapid amplification of 5′ complementary DNA ends

The 5′ UTR sequences of pre-mRNAs were identified with a SMARTer RACE 5′/3′ kit (Clontech, 634858) with primers listed in Supplemental Table S9 and Sanger sequencing. The 5′ RACE products were directionally cloned into the pRACE vector, as described in the manufacturer's protocol.

Genome information

The assembly and genome annotation were downloaded from WormBase (www.wormbase.org, version WS230 for *C. elegans*, WS227 for *C. brenneri*, WS225 for *C. briggsae* and *C. remanei*) (Harris et al. 2014). Orthologous genes were retrieved from WormBase (version WS243, ftp://ftp.wormbase.org/pub/wormbase/species/c_elegans/annotation/orthologs/). Only four-way one-to-one orthologous groups were used in this analysis. A list of essential genes was retrieved through WormMine (WS238 IM v1.2.1, http://www.wormbase.org/tools/wormmine/begin.do).

Classification of *trans*–splicing

The *trans*-splicing sites in *C. elegans* were downloaded from Allen et al. (2011). To guarantee the accurate classification of *trans*-spliced genes, *trans*-splicing sites supported by less than 10 RNA-seq reads were discarded in the following analysis. Each *trans*-splicing site was classified into one of the two categories, "SL1" or "SL2 and SL2 variants" ("SL2" for short in this study) if ≥90% of reads covering this site supported the *trans*-splicing type. Ambiguous sites with both substantial SL1 *trans*-splicing and SL2 *trans*-splicing were not considered further. If multiple *trans*-splicing sites exist on a single transcript, the *trans*-splicing type of this transcript was determined by the type of *trans*-splicing site that was covered by the largest number of reads. If a gene contained multiple mRNA isoforms, the one with the highest expression level was chosen to represent the gene. Genes without *trans*-splicing sites were defined as non-*trans*-spliced genes.

The *trans*-splicing sites of *C. brenneri*, *C. briggsae*, and *C. remanei* were parsed similarly from transcriptomic data generated in the modENCODE project (Celniker et al. 2009; Hillier et al. 2009; Gerstein et al. 2010). Specifically, transcriptomes of *C. remanei* (projects 4707–4713), *C. brenneri* (projects 4703–4706), and *C. briggsae* (projects 4495–4498 and 4693–4697) were downloaded from http://www.modencode.org. Genes annotated as SL1-only were used in the comparison between one-to-one orthologous genes without SL1 turnover.

Calculation of translational efficiency

The mRNA-seq and Ribo-seq data of four nematode species were retrieved from Stadler and Fire (2013). Genes with a normalized read count <1 in mRNA-seq were discarded to minimize the effect of inaccurate measurement of lowly expressed genes. The read count values in mRNA-seq (*mRNA*) and Ribo-seq (*Ribo*) were used to calculate ribosome density, which can infer translational initiation rate, also called translational efficiency (*TE*):

$$TE = \frac{Ribo}{mRNA}.$$

To compare translational efficiencies between orthologous genes, quantile normalization was performed on translational efficiencies among the four species.

Codon adaptation index

Synonymous codons are used with bias in the *C. elegans* genome. Following Sharp and Li (1987), we calculated the CAI, which measures the codon usage similarity of a gene to a set of 1000 most highly expressed genes in *C. elegans* (Supplemental Table S10). A detailed protocol is described in the Supplemental Methods.

Data access

The 5′ UTR sequences of pre-mRNAs of *lin-15B*, *mes-2*, and *deps-1* from this study have been submitted to GenBank (http://www.ncbi.nlm.nih.gov/genbank/) under accession numbers MF352827, MF352828, and MF352826, respectively.

Acknowledgments

The authors thank Qiwen Gan, Youli Jian, and Chonglin Yang for their help on worm experiments; Shiqin Jia from the core facility of the State Key Laboratory of Plant Genomics for her help on polysome profiling experiments; Yi Liu and Xiaofeng Cao for valuable discussions; and Bryan Moyers and John R. Speakman for critical reading of the manuscript. This work was supported by grants from the National Natural Science Foundation of China to W.Q. (31571308, 91331112), Y.-F.Y. (31601061), and Z.D. (31571535), and the "Strategic Priority Research Program" of the Chinese Academy of Sciences to Z.D. (XDB19000000). Three worm strains were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440).

Author contributions: Y.-F.Y., Q.S., and W.Q. conceived the research; Y.-F.Y., X.Z., T.Z., Z.D., and W.Q. designed the experiments; X.Z., X.M., T.Z., and Q.H. conducted the experiments; Y.-F.Y., X.Z., T.Z., Q.S., Z.D., and W.Q. analyzed the data; Y.-F.Y., X.Z., T.Z., S.W., Z.D., and W.Q. wrote the manuscript.

References

Allen MA, Hillier LW, Waterston RH, Blumenthal T. 2011. A global analysis of *C. elegans* trans-splicing. *Genome Res* **21:** 255–264.

Altelaar AF, Munoz J, Heck AJ. 2013. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* **14:** 35–48.

Andersson SG, Kurland CG. 1990. Codon preferences in free-living microorganisms. *Microbiol Rev* **54:** 198–210.

Artieri CG, Fraser HB. 2014a. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res* **24:** 2011–2021.

Artieri CG, Fraser HB. 2014b. Evolution at two levels of gene expression in yeast. *Genome Res* **24:** 411–421.

Babendure JR, Babendure JL, Ding JH, Tsien RY. 2006. Control of mammalian translation by mRNA structure near caps. *RNA* **12:** 851–861.

Bantscheff M, Lemeer S, Savitski MM, Kuster B. 2012. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **404:** 939–965.

Bao Z, Murray JI. 2011. Mounting *Caenorhabditis elegans* embryos for live imaging of embryogenesis. *Cold Spring Harb Protoc* **2011:** pdb.prot065599.

Bensimon A, Heck AJ, Aebersold R. 2012. Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem* **81:** 379–405.

Blumenthal T. 2004. Operons in eukaryotes. *Brief Funct Genomic Proteomic* **3:** 199–211.

Blumenthal T. 2005. Trans-splicing and operons. In *WormBook* (ed. The *C. elegans* Research Community), doi/10.1895/wormbook.1.5.1. http ://www.wormbook.org.

Blumenthal T, Gleason KS. 2003. *Caenorhabditis elegans* operons: form and function. *Nat Rev Genet* **4:** 112–120.

Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, et al. 2002. A global analysis of *Caenorhabditis elegans* operons. *Nature* **417:** 851–854.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129:** 897–907.

Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, MacAlpine DM, et al. 2009. Unlocking the secrets of the genome. *Nature* **459:** 927–930.

Chappell SA, Edelman GM, Mauro VP. 2006. Ribosomal tethering and clustering as mechanisms for translation initiation. *Proc Natl Acad Sci* **103:** 18077–18082.

Cheng G, Cohen L, Mikhli C, Jankowska-Anyszka M, Stepinski J, Darzynkiewicz E, Davis RE. 2007. In vivo translation and stability of *trans*-spliced mRNAs in nematode embryos. *Mol Biochem Parasitol* **153:** 95–106.

Cho S, Jin SW, Cohen A, Ellis RE. 2004. A phylogeny of *Caenorhabditis* reveals frequent loss of introns during nematode evolution. *Genome Res* **14:** 1207–1220.

Chu D, Kazana E, Bellanger N, Singh T, Tuite MF, von der Haar T. 2014. Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J* **33:** 21–34.

Conrad R, Liou RF, Blumenthal T. 1993. Functional analysis of a *C. elegans* trans-splice acceptor. *Nucleic Acids Res* **21:** 913–919.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14:** 1188–1190.

Cutter AD. 2008. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol Biol Evol* **25:** 778–786.

de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, Walther TC, Mann M. 2008. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455:** 1251–1254.

de Smit MH, van Duin J. 1990. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc Natl Acad Sci* **87:** 7668–7672.

Dvir S, Velten L, Sharon E, Zeevi D, Carey LB, Weinberger A, Segal E. 2013. Deciphering the rules by which 5′-UTR sequences affect protein expression in yeast. *Proc Natl Acad Sci* **110:** E2792–E2801.

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al. 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330:** 1775–1787.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. 2003. Global analysis of protein expression in yeast. *Nature* **425:** 737–741.

Hall MN, Gabay J, Debarbouille M, Schwartz M. 1982. A role for mRNA secondary structure in the control of translation initiation. *Nature* **295:** 616–618.

Harris TW, Baran J, Bieri T, Cabunoc A, Chan J, Chen WJ, Davis P, Done J, Grove C, Howe K, et al. 2014. WormBase 2014: new views of curated biology. *Nucleic Acids Res* **42:** D789–D793.

Hastings KEM. 2005. SL *trans*-splicing: easy come or easy go? *Trends Genet* **21:** 240–247.

Hillier LW, Miller RD, Baird SE, Chinwalla A, Fulton LA, Koboldt DC, Waterston RH. 2007. Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. *PLoS Biol* **5:** e167.

Hillier LW, Reinke V, Green P, Hirst M, Marra MA, Waterston RH. 2009. Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Res* **19:** 657–666.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324:** 218–223.

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, et al. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421:** 231–237.

Kenyon C. 1997. Environmental factors and gene activities that influence life span. In *C. elegans II* (ed. Riddle DL, et al.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Khan Z, Ford MJ, Cusanovich DA, Mitrano A, Pritchard JK, Gilad Y. 2013. Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science* **342:** 1100–1104.

Kozak M. 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15:** 8125–8148.

Krause M, Hirsh D. 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49:** 753–761.

Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324:** 255–258.

Lall S, Friedman CC, Jankowska-Anyszka M, Stepinski J, Darzynkiewicz E, Davis RE. 2004. Contribution of *trans*-splicing, 5′ -leader length, cap-poly(A) synergism, and initiation factors to nematode translation in an *Ascaris suum* embryo cell-free system. *J Biol Chem* **279:** 45573–45585.

Lasda EL, Blumenthal T. 2011. *Trans*-splicing. *Wiley Interdiscip Rev RNA* **2:** 417–434.

Lawrence J. 1999. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr Opin Genet Dev* **9:** 642–648.

Li GW, Oh E, Weissman JS. 2012. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484:** 538–541.

Li GW, Burkhardt D, Gross C, Weissman JS. 2014. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157:** 624–635.

Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6:** 26.

Maroney PA, Denker JA, Darzynkiewicz E, Laneve R, Nilsen TW. 1995. Most mRNAs in the nematode *Ascaris lumbricoides* are *trans*-spliced: a role for spliced leader addition in translational efficiency. *RNA* **1:** 714–723.

McManus CJ, May GE, Spealman P, Shteyman A. 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res* **24:** 422–430.

Nimmo R, Woollard A. 2002. Widespread organisation of *C. elegans* genes into operons: fact or function? *Bioessays* **24:** 983–987.

Paek KY, Hong KY, Ryu I, Park SM, Keum SJ, Kwon OS, Jang SK. 2015. Translation initiation mediated by RNA looping. *Proc Natl Acad Sci* **112:** 1041–1046.

Paix A, Wang Y, Smith HE, Lee CY, Calidas D, Lu T, Smith J, Schmidt H, Krause MW, Seydoux G. 2014. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in *Caenorhabditis elegans*. *Genetics* **198:** 1347–1356.

Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* **12:** 32–42.

Pop C, Rouskin S, Ingolia NT, Han L, Phizicky EM, Weissman JS, Koller D. 2014. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol* **10:** 770.

Qian W, Zhang J. 2008. Evolutionary dynamics of nematode operons: easy come, slow go. *Genome Res* **18:** 412–421.

Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. 2012. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet* **8:** e1002603.

Quax TE, Wolf YI, Koehorst JJ, Wurtzel O, van der Oost R, Ran W, Blombach F, Makarova KS, Brouns SJ, Forster AC, et al. 2013. Differential translation tunes uneven production of operon-encoded proteins. *Cell Rep* **4:** 938–944.

Saito TL, Hashimoto S, Gu SG, Morton JJ, Stadler M, Blumenthal T, Fire A, Morishita S. 2013. The transcription start site landscape of *C. elegans*. *Genome Res* **23:** 1348–1361.

Scheper GC, van der Knaap MS, Proud CG. 2007. Translation matters: protein synthesis defects in inherited disease. *Nat Rev Genet* **8:** 711–723.

Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18:** 6097–6100.
Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473:** 337–342.
Sen ND, Zhou F, Ingolia NT, Hinnebusch AG. 2015. Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res* **25:** 1196–1205.
Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. 2013. Rate-limiting steps in yeast protein translation. *Cell* **153:** 1589–1601.
Sharp PM, Li WH. 1987. The Codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15:** 1281–1295.
Spieth J, Brooke G, Kuersten S, Lea K, Blumenthal T. 1993. Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell* **73:** 521–532.
Spike CA, Bader J, Reinke V, Strome S. 2008. DEPS-1 promotes P-granule assembly and RNA interference in *C. elegans* germ cells. *Development* **135:** 983–993.
Stadler M, Fire A. 2011. Wobble base-pairing slows in vivo translation elongation in metazoans. *RNA* **17:** 2063–2073.
Stadler M, Fire A. 2013. Conserved translatome remodeling in nematode species executing a shared developmental transition. *PLoS Genet* **9:** e1003739.
Staley CA, Huang A, Nattestad M, Oshiro KT, Ray LE, Mulye T, Li ZH, Le T, Stephens JJ, Gomez SR, et al. 2012. Analysis of the 5′ untranslated region (5′UTR) of the alcohol oxidase 1 (AOX1) gene in recombinant protein expression in *Pichia pastoris*. *Gene* **496:** 118–127.
Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, et al. 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol* **1:** e45.
Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141:** 344–354.
Tyers M, Mann M. 2003. From genomics to proteomics. *Nature* **422:** 193–197.
Wallace A, Filbin ME, Veo B, McFarland C, Stepinski J, Jankowska-Anyszka M, Darzynkiewicz E, Davis RE. 2010. The nematode eukaryotic translation initiation factor 4E/G complex works with a *trans*-spliced leader stem-loop to enable efficient translation of trimethylguanosine-capped RNAs. *Mol Cell Biol* **30:** 1958–1970.
Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, et al. 2014. Mass-spectrometry-based draft of the human proteome. *Nature* **509:** 582–587.
Zaslaver A, Baugh LR, Sternberg PW. 2011. Metazoan operons accelerate recovery from growth-arrested states. *Cell* **145:** 981–992.
Zorio DA, Cheng NN, Blumenthal T, Spieth J. 1994. Operons as a common form of chromosomal organization in *C. elegans*. *Nature* **372:** 270–272.

Received November 18, 2015; accepted in revised form June 22, 2017.