The transcriptional and translational outcomes for pseudogenes in bacterial endosymbionts

Arkadiy Garber[1], Justus Nwachukwu[1,2], Ryan Stikeleather[1], Courtney York[1,2], John P. McCutcheon[1,2,*]

[1] *Biodesign Center for Mechanisms of Evolution and School of Life Sciences, Arizona State University, USA*
[2] *Howard Hughes Medical Institute, Chevy Chase, MD, USA*
[*] *Corresponding author: john.mccutcheon@asu.edu*

Intracellular bacteria in the early stages of host adaptation often show extraordinarily disrupted genomes, where up to half of their ancestral genes are found in a pseudogenized state. The mealybug *Pseudococcus longispinus* hosts two bacterial endosymbionts with high pseudogene loads, *Symbiopectobacterium endolongispinus* and *Sodalis endolongispinus*. Here, we measure the transcriptional and translational responses of these bacterial symbionts to understand how bacteria avoid (or fail to avoid) making large amounts of non-functional RNAs and proteins from these pseudogenes. Consistent with previous work, we show that pseudogenes continue to be transcribed, but at lower levels compared to intact and functional genes. Also consistent with previous work, we show that few pseudogene transcripts are translated into stable proteins. However, we find that numerous pseudogene transcripts still bind to *Symbiopectobacterium* ribosomes, and uncover a possible role for the tmRNA ribosome rescue system in the targeting of pseudogene proteins for degradation. Our results suggest a possible mechanism by which bacterial endosymbionts remove aberrant pseudogene-derived proteins during the critical time when many pseudogenes have formed but not enough time has passed for sequence evolution to erode ribosome binding sites from pseudogene transcripts.

Keywords: mealybug, pseudogene, ribosome rescue, tmRNA, translation, transcription, proteomics, ribosome profiling

Significance: Bacteria transitioning from free-living to host-dependent lifestyles often go through a transitory period where large numbers of genes are broken but not yet deleted. How cells navigate this period without producing useless or toxic gene products remains poorly understood. By combining transcriptomic, proteomic, and ribosomal profiling data from two closely related bacterial symbionts, we uncover a possible mechanism that cells use to mitigate the presence of thousands of newly formed pseudogenes. We show that pseudogenes are still widely transcribed and bind ribosomes, but are rarely translated into measurable proteins. RNA sequencing from purified ribosomes suggests that the tmRNA ribosome rescue system may act as a short-term quality control mechanism during early stages of genome reduction. These findings provide a mechanistic glimpse into how endosymbionts survive the unstable phase between gene inactivation and gene deletion, a fleeting but critical window in the evolution of endosymbiosis.

Introduction

Contrasting their metabolic, phylogenetic, and environmental diversity (Eren and Banfield, 2024), bacteria have remarkably stable and predictable genome structures. Bacterial genomes tend to retain few broken or inactivated genes (also known as pseudogenes), and encode about one functional gene every 1,000 base pairs (bp) (Kirchberger et al., 2020). Reductions from this uniform gene density—that is, fewer functional genes per region of the genome—are rare, and are mostly found in organisms that have recently undergone shifts in their environment, such as in bacteria that have recently transitioned from a free-living to an intracellular state. Living inside a eukaryotic cell brings with it strong environmental and population genetic forces that make large numbers of genes redundant (Boyd et al., 2024). While the tiniest bacterial genomes, such as those found in endosymbionts of sap-feeding insects (McCutcheon and Moran, 2011), show the typical high functional gene density, the path taken to arrive at this compact state includes transitional periods where the proportion of pseudogenes, relative to functional genes, is very high (Toh et al., 2004; Burke and Moran, 2011; Oakeson et al., 2014; Nechitaylo et al., 2021). It is now known that in the early stages of endosymbiosis, when a free-living bacterium transitions to becoming a vertically transmitted symbiont, large numbers of pseudogenes can accumulate. This happens due to functional redundancy with the host cell and other symbionts (Moran et al., 2009; Koga and Moran, 2014), selection against genes that promote a host immune response (Amiel et al., 2010), and a genome-wide reduction in the efficacy of purifying selection (Lerat and Ochman, 2005). In genomes from bacteria that have been caught in a recent transition to the endosymbiotic state, the number of pseudogenes can rival the number of intact genes (Toh et al., 2004; Oakeson et al., 2014). Large numbers of pseudogenes have also been observed in several intracellular bacterial pathogens (Benjak et al., 2017; Cole et al., 2001), likely for similar reasons to those described above for host-beneficial endosymbionts.

How do bacterial cells deal with having hundreds, or even thousands, of broken genes on their genomes? It is reasonable to guess that producing non-functional mRNA

Table 1: Summary of pseudogenes and coding densities across three bacterial genomes studied here. The numbers in parentheses for the Intact genes and pseudogenes columns represent the amount of genome taken up by these annotations. Mb = megabase pairs, Kb = kilobase pairs. *Intergenic DNA excluding pseudogenes. **Coding density calculations exclude non-protein-coding genes.

Bacterium	**Genome size**	**Intact genes**	**Pseudogenes**	**Non-coding DNA***	**Coding density****
Sodalis praecaptivus	5.159 Mb	4331 (4.1 Mb)	90 (103 Kb)	929 Kb	83.3%
Symbiopectobacterium endolongispinus	4.492 Mb	2559 (2.0 Mb)	2679 (1.8 Mb)	678 Kb	44.5%
Sodalis endolongispinus	3.728 Mb	2294 (1.9 Mb)	1807 (1.1 Mb)	730 Kb	51.0 %

would be less problematic than producing non-functional protein products, for both energetic (Lynch and Marinov, 2015) and mechanistic (Kuo and Ochman, 2010) reasons. Previous studies on the transcriptomic and proteomic response to bacterial pseudogene expression support this hypothesis. The tsetse fly endosymbiont *Sodalis glossinidius* shows widespread transcription of its pseudogenes, which number in the thousands (Goodhead et al., 2020). Similar patterns of extensive production of mRNA from pseudogenes were observed in the leprosy bacillus *Mycobacterium leprae* (Suzuki et al., 2006; Williams et al., 2009) and the insect symbiont *Candidatus* Streptomyces philanthi (Nechitaylo et al., 2021). However, most evidence suggests that transcripts produced from pseudogenes do not often lead to the production of measurable protein products. In *M. leprae*, many transcribed pseudogenes lack ribosomal binding sites or start codons, which should reduce the efficiency with which they initiate translation (Williams et al., 2009).

In the insect symbionts *Sodalis glossinidius* and *Candidatus* Streptomyces philanthi, proteomics revealed that relatively few of the numerous pseudogenes found in their transcriptomes were translated into protein products (Goodhead et al., 2020; Nechitaylo et al., 2021). However, a survey in *Salmonella enterica* found that about two-thirds of its roughly 160 pseudogenes produced detectable peptides, indicating successful (if often partial) translation of ostensibly inactivated genes (Feng et al., 2022). In *Mycobacterium tuberculosis*, which carries fewer pseudogenes than its leprosy-causing relative, ribosome profiling revealed that numerous pseudogene-derived transcripts are being actively translated (Smith et al., 2022), suggesting that, in some cases, bacteria may not be able to completely stop the translation of non-functional transcripts, regardless of their downstream coding potential.

Here, we use the endosymbionts of the long-tailed mealybug, *Pseudococcus longispinus*, to better understand transcriptional and translational outcomes when large numbers of pseudogenes are present on a bacterial genome. *P. longispinus* has an unusual, nested symbiont structure, where two species of recently acquired endosymbionts, *Candidatus* Sodalis endolongispinus (hereafter, *Sodalis endo.*) and *Candidatus* Symbiopectobacterium endolongispinus (hereafter, *Symbiopectobacterium endo.*), reside within the cytoplasm of a long-established endosymbiont called *Candidatus* Tremblaya princeps (hereafter, *Tremblaya princeps*; ,k and McCutcheon 2016; Garber et al., 2021). Both *Sodalis endo.* and *Symbiopectobacterium endo.* have large genomes containing thousands of pseudogenes, and so are useful models to study the transcriptional and translational effects of widespread pseudogene accumulation.

RESULTS

Symbiopectobacterium endo. maintains a greater load of younger, more recently formed, pseudogenes

The genomes of *Sodalis endo.* and *Symbiopectobacterium endo.* are 3.7 Mb and 4.5 Mb, respectively (Garber et al., 2021). Both genomes have many pseudogenes and appear to be in the early stages of genome erosion (**Table 1**). *Sodalis endo.* is closely related to the free-living bacteria *Sodalis praecaptivus* HS1 (Clayton et al., 2012; Chari et al., 2015), with which it shares about 90% genome-wide average nucleotide identity (ANI). Assuming an ancestral genome size similar to *S. praecaptivus*, the symbiont *Sodalis endo.* has lost about 30% of its genome. *Symbiopectobacterium endo.* has no close genus-level relatives that are free-living, but is phylogenetically within the *Brenneria*/*Pectobacterium* clade. The *Symbiopectobacterium* genome is close in length to the average 5 Mb-sized genomes seen across this clade,

Table 2: Broad categories assigned to pseudogenes in young endosymbionts from *P. longispinus*. The most numerous pseudogenes in both endosymbionts and *S. praecaptivus* are of the truncated type, and about 19% of pseudogenes are cryptic in the two symbionts *Symbiopectobacterium* and *Sodalis*. Near-complete pseudogenes, generally considered as the youngest (most recently formed) pseudogenes comprise a third (33%) of all pseudogenes in *Symbiopectobacterium*, but only 16% in *Sodalis endo.* and 4% in *S. praecaptivus*.

Pseudogene category	**Number in *Symbiopectobacterium***	**Number in *Sodalis endo.***	**Number in *S. praecaptivus***
Cryptic	499 (19%)	331 (19%)	27 (14%)
Truncated	1285 (48%)	1173 (65%)	161 (82%)
Near-complete	895 (33%)	303 (16%)	8 (4%)

suggesting that it may have recently diverged from a free-living bacterium.

While both symbionts have similarly low coding densities of ~45-50%, they differ in the types of pseudogenes they carry on their genomes (**Table 2**). We divided symbiont pseudogenes into three broad (and sometimes overlapping) categories: near-complete, truncated, and cryptic. Near-complete pseudogenes are very similar in size and structure to homologous functional genes on other genomes, but contain one or more indels or in-frame stop codons that disrupt the ancestral reading frame into two or more open reading frames (ORFs). Truncated pseudogenes are those where less than 65% of the ancestral gene length remains on the genome. Cryptic pseudogenes are those that retain few features of a functional gene and are difficult to detect by means other than sensitive ORF-independent sequence alignment algorithms. In both symbionts, the most common form of pseudogene is of the truncated type (**Table 2**).

We assume that small sequence changes—a small frameshift-inducing insertion or deletion, or a single-base substitution that changes a sense codon to a nonsense codon—occur more frequently than larger ones, and are thus more likely to initially break a gene (Liu et al., 2004; Kuo and Ochman, 2010; Daneels et al., 2018). As newly broken genes accumulate more deletions over time, they eventually appear as partial remnants of genes with significant portions of the original ORF missing or unrecognizable because of lack of purifying selection to maintain the coding sequence. Following this assumption, we find that in comparison with *Sodalis endo*., *Symbiopectobacterium endo*. has a greater density of the (assumed) younger near-complete pseudogenes (**Table 2**). While this suggests that *Sodalis endo*. is an older endosymbiont, other mechanistic factors may be at play, such as the presence of numerous active transposases on the *Sodalis endo*. genome (Garber et al., 2021), which can catalyze deletions, including partial gene deletions (Plague et al., 2008), resulting in faster clearance of near-complete pseudogenes. Nevertheless, overall, it appears that *Symbiopectobacterium endo*. has a larger fraction of the more recently formed, or younger, class of near-complete pseudogenes.

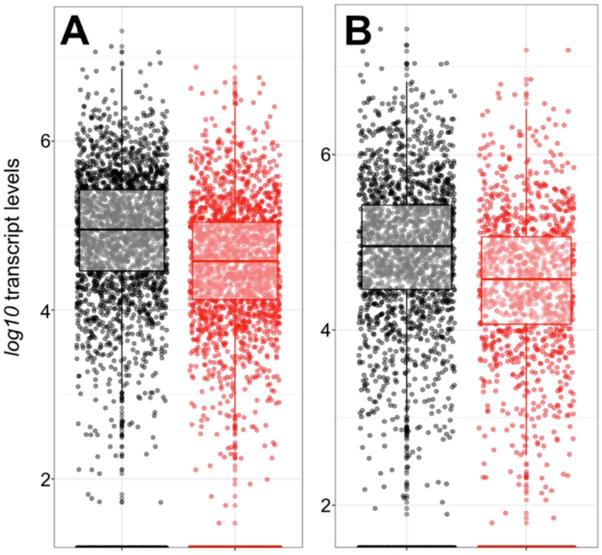


Figure 1: Gene expression levels in A) *Symbiopectobacterium endo*., B) *Sodalis endo*. Red dots on the right side of each plot indicate the pseudogenes on each genome.

To understand the downstream molecular consequences of a genome encoding large numbers of pseudogenes, we sequenced whole transcriptomes from mealybug bacteriomes (the bacteria-housing organs) to measure transcription across these endosymbiont genomes.

Pseudogenes are widely transcribed without a significant decrease in expression level

RNA sequencing reveals that pseudogenes continue to be transcribed, but, on average, at lower levels than intact genes. Out of 2679 pseudogenes encoded on the *Symbiopectobacterium endo*. genome, we detect 1978 (74%) within its transcriptome. We detect transcription of 2278 of 2559 (89%) of intact genes from *Symbiopectobacterium endo*. However, despite comparable numbers of intact genes (2278) and pseudogenes (1978) in the transcriptome of *Symbiopectobacterium endo*., pseudogenes are expressed at significantly lower levels compared to intact genes (**Figure 1A**) making up only 28% of the transcriptome by RNA read abundance, measured with normalized transcripts per million (TPM). This is also the case in *Sodalis endo*., where pseudogenes are not only expressed at lower levels (**Figure 1B**), but are also less likely to be transcribed. We detected 1152 of 1807 (64%) pseudogenes from *Sodalis endo*., which are also less abundant in the transcriptome, making up 19% by TPM-normalized read abundance.

The lowered abundance of pseudogenes in both symbionts' transcriptome suggests that pseudogene transcripts may be inhibited or downregulated in some way, or, in pseudogenes that have been around longer, the RNA polymerase binding sites may have eroded away on the genome. However, it also could be that pseudogenes may simply be more likely to arise from lowly expressed genes. Given the close relationship between the free-living *Sodalis praecaptivus* and *Sodalis endo*., we had the unique opportunity to measure the expression levels of gene homologs that were pseudogenized on an endosymbiont genome but intact on a genome from a free-living close relative (**Figure 2A**). Of the 2128 orthologs shared between these two *Sodalis* genomes, 677 are pseudogenes in *Sodalis endo*. but remain intact in *S. praecaptivus* (**Figure 2B**). In *S. praecaptivus*, these 677 genes are expressed at lower levels compared to genes that appear intact in *Sodalis endo*. (**Figure 2C**), supporting the hypothesis that lowly expressed genes are more likely to become pseudogenes. This pattern is consistent with previous work showing that highly expressed genes experience stronger purifying selection than lowly expressed genes (Yannai et al., 2018; Roberts

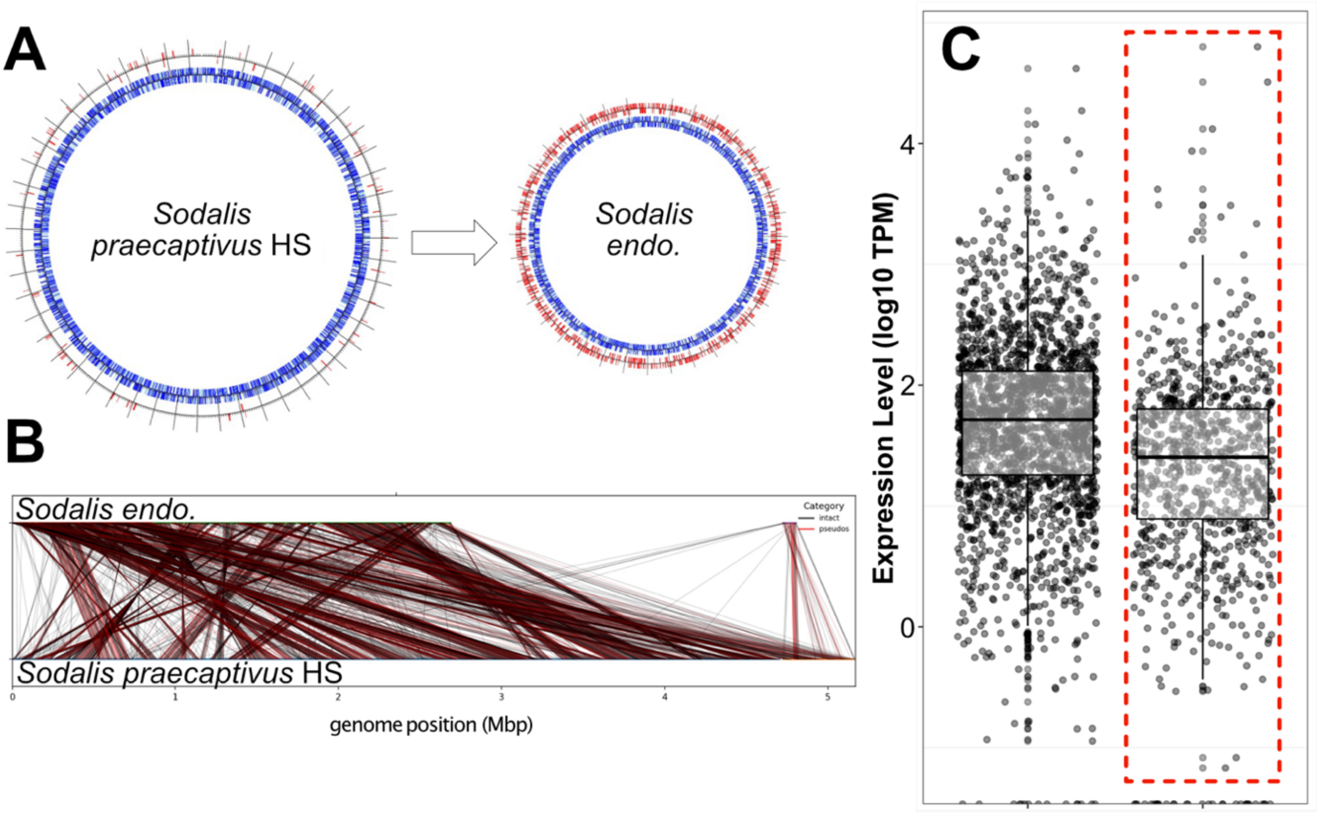


Figure 2: A) Circular genome diagrams of free-living *S. praecaptivus* and *Sodalis endo*. showing predicted gene regions as blue radial lines along the inner track. Pseudogenes are visualized on the outer track as red radial lines. B) Genome synteny plot showing relationship of genes on the genome of *Sodalis endo*. with orthologs from *S. praecaptivus*. Lines colored red indicate genes that become pseudogenized, while black lines indicate genes that remain intact on Sodalis endo. C) Gene expression from *Sodalis praecaptivus*, showing a comparison between genes that remain intact and those that become pseudogenized on *Sodalis endo*, which are enclosed in a red-dotted box.

and Josephs, 2023). It is also possible that highly expressed genes that become pseudogenes have a higher likelihood of causing negative impacts downstream, so are more quickly removed via selection (Kuo and Ochman, 2010). Nevertheless, our data here suggest that what looks like lowered expression of pseudogenes compared to intact genes may simply reflect the fact that lowly expressed genes are more likely to become pseudogenes compared with highly expressed genes.

Having shown that a significant amount of pseudogene expression still occurs in these two endosymbionts, it seemed possible that these transcripts could still bind to ribosomes and attempt translation, with potentially deleterious downstream consequences. To explore this possibility, we next sequenced RNA transcripts that were bound to ribosomes using a modified ribosomal profiling method for bacteria (Ingolia et al., 2012; Mohammad and Buskirk, 2019).

Pseudogene transcripts continue to bind ribosomes in *Symbiopectobacterium endo.*

Previous studies have found persistent transcription of pseudogenes combined with comparatively little protein products resulting from these broken genes (Goodhead et al., 2020; Nechitaylo et al., 2021; Feng et al., 2022). These patterns suggest that one or more post-transcriptional mechanisms are acting to prevent large amounts of bad protein from being produced. In search of possible mechanisms preventing the accumulation of large amounts of pseudogene-derived proteins, we isolated ribosomes from *P. longispinus* using sucrose gradient ultracentrifugation and sequenced all RNA that co-purifies with ribosome fractions. Our assumption here is that mRNA that co-purifies with ribosomes is likely to be actively translated, allowing us to infer the translational activities of *Sodalis endo*. and *Symbiopectobacterium endo.*

Ribosome-bound transcripts from *Symbiopectobacterium endo*. ribosomes dominated our sequencing efforts, yielding on average 30x more reads compared to *Sodalis endo*. (**Table 3**). The dominance of *Symbiopectobacterium endo*. RNA from ribosome-bound samples is inconsistent with previous measurements of the relative abundance of the two endosymbionts as assayed by genome coverage and RNA-FISH microscopy imaging (Garber et al., 2021), as well as the whole-transcriptome RNA sequencing and proteomic experiments performed here (**Table 3**). We attempted our total RNA isolations both with and without the translation inhibiting antibiotic chloramphenicol, and in both experiments obtained similar imbalances favoring

Table 3: Relative abundances of various biomolecules in the two intra-*Tremblaya* endosymbionts. RNA-FISH is a proxy for SSU rRNA; genomic read coverage measures the relative number of genome copies, whole-transcriptome mRNA measures the total amount of transcripts, while ribosome-bound mRNA measures the number of ribosome-bound transcripts undergoing translation; proteome abundance is measured as a proxy of peptide peak abundance from the initial mass spectrometry scan (MS1).

Symbiont	RNA-FISH	genomic read coverage	whole-transcriptome mRNA	ribosome-bound mRNA	proteome (MS1 peak abundance)
Sodalis endo.	29%	35%	27%	3%	39%
Symbiopectobacterium endo.	71%	65%	73%	97%	61%

Symbiopectobacterium endo. We suspect that *Sodalis endo*. ribosomes are less stable than *Symbiopectobacterium endo*. ribosomes in our purification protocol. Unfortunately, we do not know if this is an experimental artifact (*Sodalis endo*. ribosomes are less stable in the buffers we use) or whether it reflects actual biology (*Sodalis endo*. ribosomes are present but are less likely to be in active 70S conformations). While we report below on the RNAs bound to *Sodalis endo*. ribosomes, we caution that these results may be unreliable or biased in some way that we cannot predict.

We find further contrast between the two symbionts in the proportions of pseudogene transcripts bound to their ribosomes. *Symbiopectobacterium* attempts translation of nearly all of its pseudogene transcripts, with similar counts of intact (2202) and pseudogene (1923) transcripts bound to ribosomes (orange dots in **Figure 3A**). By contrast, only 151 out of a total 786 detected pseudogene transcripts were found among the ribosome-bound RNA in *Sodalis endo*. (orange dots in **Figure 3B**). In both endosymbionts, ribosome-bound RNA from pseudogene transcripts is less abundant compared with intact gene transcripts (22% in *Symbiopectobacterium endo*. and 8% *Sodalis endo*.), suggesting that transcripts encoding pseudogenes are less likely to stably bind ribosomes or are more likely to terminate prematurely, or both (**Supplemental Figure 1**).

Surprisingly, nearly 40% of all RNA obtained from *Symbiopectobacterium endo.* ribosomes is from the transfer-messenger RNA (tmRNA; **Figure 3C**). This small RNA, which contains both an mRNA and a tRNA-like domain (Keiler et al., 1996), is involved in recycling bacterial ribosomes that have stalled on broken mRNA transcripts (Moore and Sauer 2005; Janssen and Hayes, 2012; Keiler 2008). The tmRNA gene is conserved across bacteria, and transcripts from it are often the most abundant and stable RNAs in the cell (Ray and Apirion, 1979). In *Symbiopectobacterium endo*., tmRNA comprises about 1% of the total transcriptomic RNA, and so its presence as 40% of all ribosome-bound transcripts is significant. In both *Sodalis endo*. and the free-living *Sodalis praecaptivus*, the enrichment of tmRNA on ribosomes vs. total transcriptomes is less pronounced (3.7% vs. 4.1% in *Sodalis endo*. and 10% vs. 7.4% in *Sodalis praecaptivus*, **Figure 3D-E**).

The large numbers of pseudogene transcripts binding to ribosomes, particularly in *Symbiopectobacterium endo*, raises the possibility that significant levels of non-functional or toxic protein products could accumulate in these cells. To investigate whether or not proteins were produced from these ribosome-bound pseudogene transcripts, we performed shotgun mass spectrometry (MS) proteomics on mealybug bacteriomes.

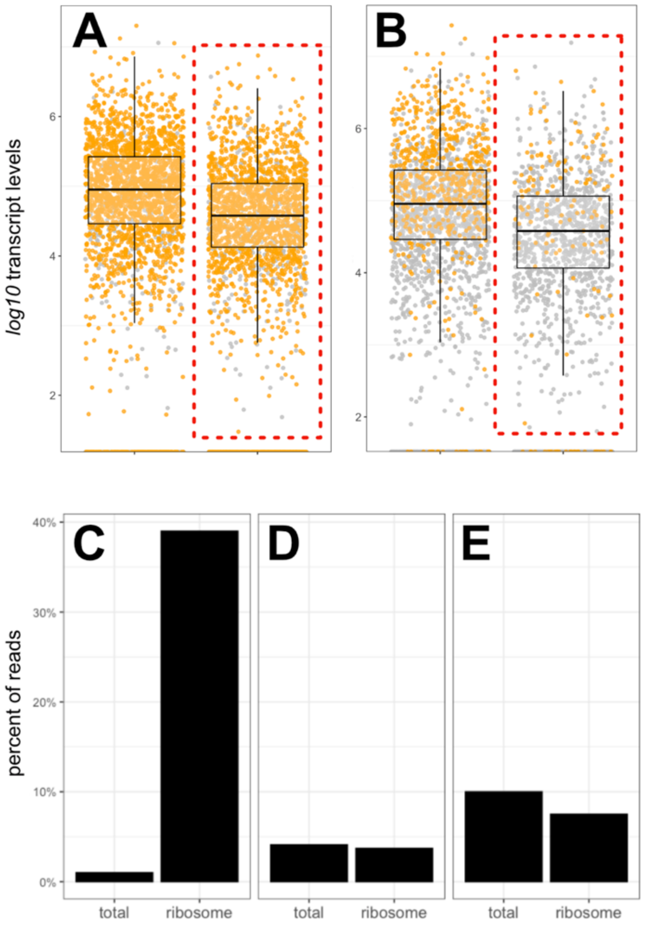


Figure 3: Gene expression levels in A) *Symbiopectobacterium endo*. and B) *Sodalis endo* (same data as in Figure 1). Pseudogenes are highlighted on the right side of each plot with red dotted boxes. Dots filled in with yellow indicate genes identified among the ribosome-bound RNA. Panels C-E show the relative amounts of tmRNA among the total and ribosome-bound transcriptomes in C) *Symbiopectobacterium endo*., D) *Sodalis endo*, and E) *Sodalis praecaptivus*.

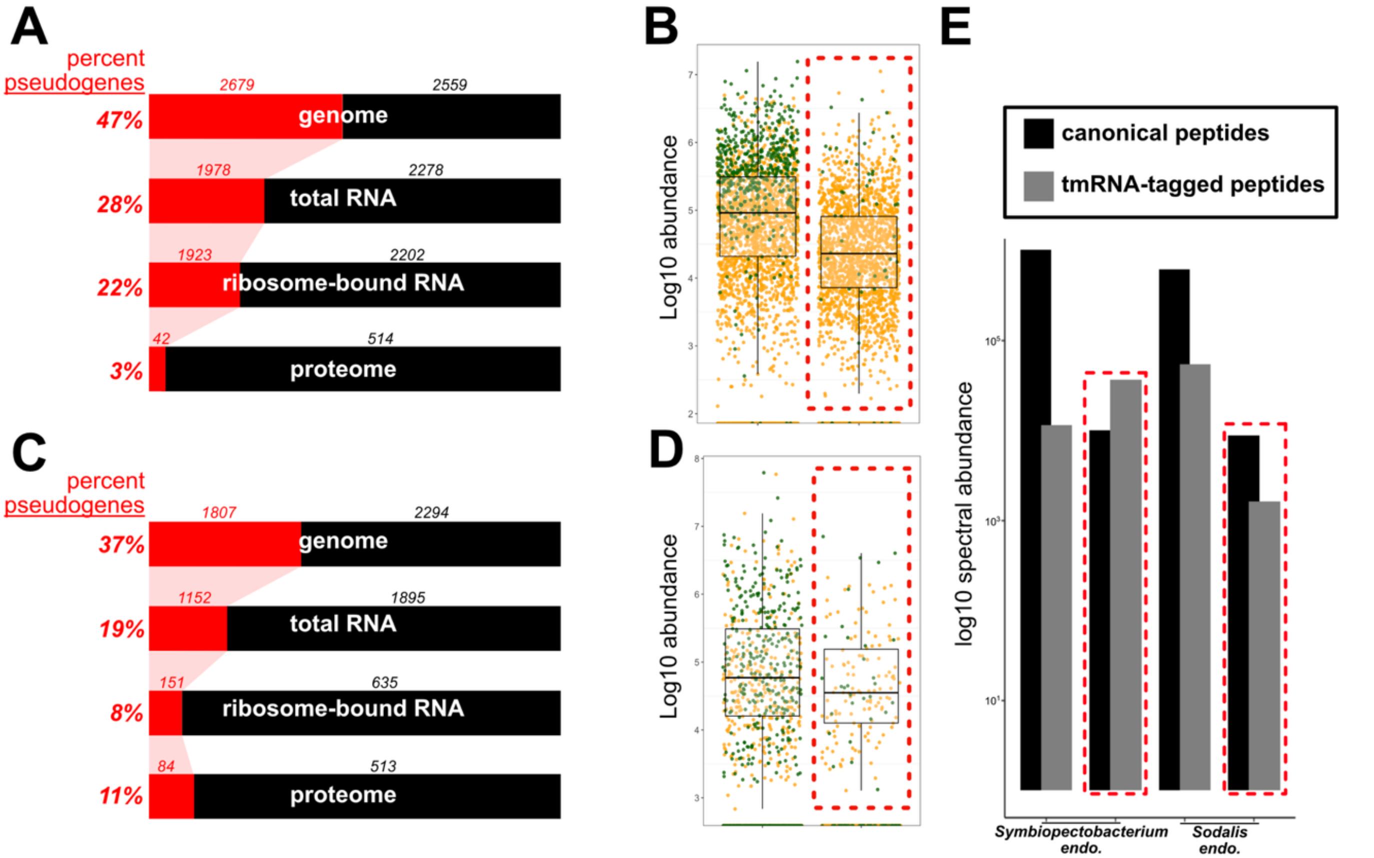


Figure 4: A) Sankey diagram showing the relative proportion of each biomolecule that corresponds to pseudogenes (shown in red) in Symbiopectobacterium. For genome, percent pseudogene refers to proportion of bases covered by pseudogenes. For the total and ribosome-bound RNA, percent pseudogene corresponds to percent of total RNA abundance. For proteome, percent pseudogene corresponds to percent of total peptides by abundance. Small numbers above each Sankey bar indicate the numbers of genes or pseudogenes identified at each stage. B) Translation levels in *Symbiopectobacterium endo*., with pseudogenes denoted in red-dotted boxes. C) Sankey diagram for Sodalis endo. D) Translation levels in *Sodalis endo*., with pseudogenes in red-dotted boxes E) Relative peptide abundance of canonical and tmRNA-tagged peptides in *Sodalis* and *Symbiopectobacterium endo*. Spectral abundance values are shown on a log-scale, with arbitrary spectral units obtained from the mass spectrometer. Pseudogenes shown in red-dotted boxes.

Protein products from pseudogenes are rare, particularly in *Symbiopectobacterium endo.*

MS proteomics of mealybug bacteriomes yielded 6041 proteins from the host mealybug, which comprised the vast majority of our MS data. The rest of the proteome is split roughly evenly between *Tremblaya* (120 proteins, 6% of peptides by abundance), *Sodalis endo*. (597 proteins, 5% by abundance) and *Symbiopectobacterium endo*. (558 proteins, 4% by abundance). Among the 558 proteins detected from *Symbiopectobacterium endo*., only 42 are pseudogenes (**Table 4**). Based on the relative abundance of each identified peptide measured via intensity of MS1 precursor ions (Palomba et al., 2021), pseudogene-derived peptides are in low abundance, making up less than 3% of the total *Symbiopectobacterium endo*. proteome (**Figure 4A**). The majority of the *Symbiopectobacterium endo*. proteome consists of peptides derived from intact transcripts that are abundant among isolated ribosomes (**Figure 4B**). In *Sodalis endo*., by contrast, we find little relationship between ribosome-bound transcript levels and protein presence, with pseudogene-derived peptides comprising 11% of the proteome by abundance (**Figure 4C-D**).

Given the high abundance of tmRNA bound to *Symbiopectobacterium endo.* ribosomes, we looked for potential products of ribosome rescue in the proteome. The coding region of tmRNA codes for a short peptide tag that is appended to proteins stalled on broken transcripts, targeting them for degradation (Keiler, 2008). When we included peptides with this tag in our proteomic search space, we identified additional peptides bearing the ANDENYALAA tag in *Symbiopectobacterium endo.* and the ANDSQFESKTALAA tag in *Sodalis endo*. These tmRNA-tagged peptides are in relatively low abundance compared to unmodified peptides, except from *Symbiopectobacterium endo*. pseudogenes (**Figure 4E**). Based on relative spectral abundances, pseudogene-derived

Table 4: Summary of pseudogenes across the symbionts' genomes, transcriptomes, and ribosome-bound RNA. tmRNA abundance is shown as a percentage of total mRNA translation (excluding tRNAs, rRNAs, and other non-coding RNA sequences).

Organism	Pseudogenes encoded	Pseudogenes transcribed	Pseudogenes with ribosome-bound transcripts	tmRNA percent in ribosomes	Proteins from pseudogenes detected
Symbiopecto endo.	2679	1978	1923	39.3%	42
Sodalis endo.	1807	1152	151	3.7%	84

peptides with the tmRNA tag outnumber unmodified pseudogene peptides 4-fold in *Symbiopectobacterium endo*. Overall, these proteomic data suggest a higher level of tmRNA activity in *Symbiopectobacterium endo.*, consistent with the large amount of tmRNA that we find co-purifying with its ribosomes (**Figure 3C**).

Discussion

Short-term vs. long-term adaptations to large numbers of pseudogenes

When a pseudogene is first formed, the upstream sequence elements involved in RNA polymerase and ribosome recognition are unchanged, and therefore transcription and translation should attempt to proceed as normal. Depending on the nature of the brand-new pseudogene, the fitness effect of continued transcription and translation might vary from minor, such as a deletion that affects the very end of a lowly expressed gene, to major, such as a deletion that alters the reading frame near the middle of a highly expressed gene. Over evolutionary time, selection will eliminate the sequence elements of this pseudogene that allow it to be transcribed and translated. We see evidence of this type of evolution in the global analysis of start codons and Shine-Dalgarno (ribosome-binding) sites of pseudogenes vs. functional genes: pseudogenes are less likely to have upstream Shine-Dalgarno sites (**Supplemental Figure 3**) and are also less likely to start with the preferred ATG initiation codon (Belinky et al., 2017) (**Supplemental Figure 4**). Of special interest to us in this study was to uncover any short-term mechanistic responses that bacteria might use to repress transcription and translation of pseudogenes over short time frames, before evolution has had time to eliminate the RNA polymerase and ribosome binding sites. For this reason, we were especially interested in the mechanisms that *Symbiopectobacterium endo*. might use to prevent the formation of aberrant proteins, because our analysis of pseudogene age suggests that this endosymbiont carries a significant fraction of newer pseudogenes on its genome (**Table 2**).

Lowly expressed genes are more likely to become pseudogenes

We reasoned that there might not be a strong mechanistic response to pseudogene formation at the level of transcription. While some mechanisms are known to affect mRNA stability in bacteria (Mohanty and Kushner, 2016), we know of no mechanism in bacteria similar to nonsense-mediated decay in eukaryotes (Chang et al., 2007) that might alter the stability of pseudogene transcripts specifically. When looking at the levels of transcription between functional genes and pseudogenes globally, we, like others before us (Goodhead et al., 2020; Nechitaylo et al., 2021; Feng et al., 2022), see an average decrease in the transcription levels of pseudogenes (**Figure 1A-B**). Some of this reduced expression is likely a consequence of the loss of transcription binding sites due to sequence evolution of older pseudogenes. However, we wondered how much of this signal was simply due to the higher probability of lowly expressed genes becoming pseudogenes, rather than some global mechanistic response that would reduce pseudogene expression. The idea here is that, on average, lowly transcribed genes are generally under weaker purifying selection (Yannai et al., 2018; Roberts and Josephs, 2023) and therefore, might be less likely to be important in the context of endosymbiosis. To test this hypothesis, we compared transcript abundance of *Sodalis endo.* pseudogenes to their intact orthologs from *Sodalis praecaptivus*. We found that genes destined to become pseudogenes in *Sodalis endo.* are lower in abundance within the transcriptome of *Sodalis praecaptivus* (**Figure 2C**).

Along with the fact that no known mechanism for systematically down-regulating pseudogene transcripts in bacteria exists, these results suggest that there is no transcriptional response to becoming a pseudogene on short time scales: transcription continues as normal until sequence evolution alters transcriptional binding sites. Weaker transcription of pseudogenes in *Sodalis endo*. compared to *Symbiopectobacterium endo*. (**Table 4**) may thus reflect a difference in how long each symbiont has been evolving under weakened selection, with *Symbiopectobacterium endo*. representing the younger endosymbiont, captured before most of the degradation of pseudogene transcriptional signals has had time to occur.

Transcripts from nascent pseudogenes bind ribosomes but rarely make protein products

Before sequence evolution has time to change the promoter sequences of a pseudogene, we expect it to be transcribed as usual and for the transcripts to bind ribosomes as usual. We see evidence of this in our purified ribosome preparations, where a substantial amount of *Symbiopectobacterium endo*. pseudogene transcripts are found bound to ribosomes

(**Figure 3B**). However, very few of those transcripts are made into protein products, or very few of these protein products remain stable or abundant enough in the cell for us to measure in mass spectrometry experiments (**Figure 4A**). What mechanism might be at work to prevent pseudogene transcripts from being made into aberrant protein products, before evolution has had time to erode away the ribosome binding signals? The high amounts of tmRNA from *Symbiopectobacterium endo.* in our purified ribosome preparations suggests a possible mechanism. The tmRNA is known to be abundant in bacterial RNA-seq datasets (Engelhardt et al., 2020), but we were unsure how abundant it was in ribosome profiling experiments. We analyzed ribosomal profiling data from *Escherichia coli* (Mangano et al., 2022), as well as *Streptomyces venezuelae* and *Streptomyces griseus* (Kim et al., 2020), which were generated using methods similar to ours (direct sequencing of ribosome-bound RNA without micrococcal nuclease digestion). The percent of reads mapping to tmRNA ranged from 0.02% in *S. griseus*, to 0.44% in *S. venezuelae*, to 3.4% in *E. coli* (**Supplemental Table 1**). Other studies using ribosome footprinting showed similar proportions of tmRNA, never exceeding 4.7% of total ribosome-protected RNA fragments (**Supplemental Figure 5**). These data suggest that our measurement of over 39% of reads from *Symbiopectobacterium* being from tmRNA is significant (**Figure 3D, Table 4**).

The function of tmRNA is to rescue stalled ribosomes that have attempted translation on a broken mRNA transcript. The tmRNA first enters the ribosomal aminoacyl-tRNA binding site via its tRNA-like domain, causing the ribosome to switch over to the templated mRNA sequence of tmRNA and to translate a short 10-peptide long chain that is appended to the nascent protein (Moore and Sauer, 2005). This mechanism is colloquially known as ribosome rescue and is thought to allow bacteria to quickly recycle inefficient or stalled ribosomes (Moore and Sauer, 2005). We found evidence for products of active ribosomal rescue (protein fragments appended with the short peptide encoded on the tmRNA CDS region) in our mass spectrometry data, in particular in *Symbiopectobacterium endo*., where there are about 4-fold more pseudogene peptides with the tmRNA tag than without (comprising 78% of pseudogene-derived peptides). This is in contrast to products of *Symbiopectobacterium* intact genes, where only one percent of peptides are found with a tmRNA tag. By contrast, in *Sodalis endo*., rescue-derived products are in the minority among peptides from both intact and pseudogenes, but still enriched among the pseudogene products.

Proposed mechanism for silencing of newly formed pseudogenes

Our finding that large amounts of tmRNA are bound to *Symbiopectobacterium endo.* ribosomes suggests a role for this RNA in eliminating aberrant pseudogene-derived proteins from the proteome. However, tmRNA only acts on broken mRNA transcripts, and so at least one more step is required for our model to work. It has been shown that ribosome collisions can recruit the activity of tmRNA due the cleavage of mRNA on collided ribosomes upstream of the stall site by the protein SmrB, which then allows tmRNA and its partner protein SmpB to enter the ribosome and terminate translation (Saito et al., 2022). Both the SmrB and SmpB proteins and their transcripts were detected in our mass spectrometry and transcriptomic data (**Supplemental Figure 8**). It seems reasonable to hypothesize that pseudogenes might experience ribosome collisions more frequently, due to shifts in codon bias that occur when an open reading frame is perturbed. This subtle consequence of codon usage may have a significant impact on translation efficiency and stalling due to higher incidences of rare codons and cognate tRNA scarcity (Roche and Sauer, 1999; Samatova et al., 2021).

Overall, our data suggest that bacteria do little to prevent the transcription and translation of pseudogenes as they first emerge. Over time, sequence evolution away from preferred RNA polymerase and ribosome binding sites will decrease or prevent the transcription and translation of pseudogenes. How bacteria prevent newly formed pseudogenes from producing products has been less clear. Here, we present data suggesting the ribosome rescue system as a mechanism to both rapidly degrade proteins made from pseudogenes and rescue ribosomes that are pausing, stalling, and colliding on non-optimal sets of codons. This ribosome rescue system may be useful for bacteria that have found themselves in a state where lots of pseudogenes exist on their genome but the time needed to eliminate ribosomal initiation sites through sequence evolution has not yet passed.

MATERIALS AND METHODS

Insect rearing and bacterial culturing

Pseudococcus longispinus populations were reared on sprouted potatoes at 25°C, 77% relative humidity, and a 12h light/dark cycle in a Percival 136LL incubator. *Sodalis praecaptivus* HS was grown on liquid lysogeny broth (LB, Miller) at 30 ˚C with vigorous shaking.

RNA extraction and sequencing of mealybugs and *Sodalis praecaptivus*

For transcriptomic sequencing, *P. longispinus* mealybug bacteriomes were dissected, immediately flash frozen in liquid nitrogen, and then crushed with a plastic pestle inside sterile microcentrifuge tubes (1.5 mL). Tissue lysates were clarified (to remove insoluble cell debris) with a 30 min spin at 20,000 g (4°C). RNA from clarified lysates was purified using the Qiagen RNeasy kit with 1 mM DTT added to the lysis buffer (Buffer RLT), along with 0.0002 U/l RNase Inhibitor (Thermo Fisher Scientific). We performed DNA digestion with DNase I prior to washing and elution with double-distilled H2O (ddH2O). Purified RNA quality was assessed using the Agilent 4200 TapeStation. Depletion of ribosomal RNA was performed with Illumina's RiboZero kit using standard probes. Sequencing was performed by Genewiz with the Illumina NextSeq 500 platform.

Sodalis praecaptivus HS cells were harvested at mid-log phase of growth (OD600 between 0.6 and 0.9) with 15 min of centrifugation at 5000 g (4°C). Cells were lysed by flash-freezing in liquid nitrogen and passing through a 27-gauge syringe 15-20 times. Lysates were clarified as discussed above and RNA extracted following the same protocol as mealybug samples.

Ribosomal profiling of mealybugs and *Sodalis praecaptivus*

We performed ribosomal profiling on whole mealybugs, as well as on dissected bacteriome tissue. In both cases we used ice-cold Buffer A (20 mM Tris-HCl pH 7.5, 100 mM NH4Cl, 10.5 mM MgOAC, 0.5 mM EDTA, 0.5 M chloramphenicol), containing 0.0002 U/l RNase Inhibitor, 0.0002 U/l DNase I, and two tablets of Pierce Protease Inhibitor (Thermo Fisher Scientific) for lysis. Additionally, in the bacteriome preparation, we included 25 µg/mL chloramphenicol. Samples were homogenized using a Dounce homogenizer with 15-20 mechanical strokes. The resulting lysates were centrifuged at 20,000 g for 30 minutes (4°C). Pellets were discarded and supernatant (1 ml each) layered onto 12 mL 10-50% sucrose gradients made with Buffer B (20 mM Tris-HCl pH 7.5, 500 mM NH4Cl, 10.5 mM MgOAC, 0.5 mM EDTA, 1.1 M sucrose). We also added 2-mercaptoethanol (420 µl/L), benzamidine (BME, 1 ml/L of 100 mM), and phenylmethylsulfonyl fluoride (PMSF, 2 mL/L 50 mM) to both Buffer A and B immediately before use. The gradient was prepared with a BioComp Gradient Master, and tubes inserted into the Swinging Bucket 40 Ti rotor and spun for 2.5 h at 200,000 g, 4°C. Resulting gradients were cut using the BioComp Piston Gradient Fractionator, with elution absorbances measured at 260 nm to detect nucleic acids. Ribosomal fractions corresponding to 70S ribosomes and polysomes were taken for RNA extraction and sequencing (**Supplemental Figure 9**).

S. praecaptivus HS colonies were picked from LB agar plates and sub-cultured into conical flasks containing LB. Cells were harvested in the mid-log phase of growth and pelleted with centrifugation at 5000 g for 15 min (4°C) and homogenized using a high-pressure homogenizer (5 rounds, 20 psi). Homogenization was performed using Buffer A, with 0.0002 U/l RNase Inhibitor, 0.0002 U/l DNase I, two tablets of Pierce Protease Inhibitor, and 25 µg/mL chloramphenicol. The resulting lysates were clarified and subjected to sucrose-gradient ultracentrifugation as described above for mealybug samples. Elution absorbances measured at 260 nm yielded monosome peaks starting at fraction 10 (roughly halfway down in the gradient), and more clearly-defined polysome peaks in fractions 14-20 (**Supplemental Figure 10**).

RNA was extracted using TRIzol (Thermo Fisher Scientific) with 20% chloroform. Each fraction from the sucrose gradient (600 µl) was combined with 2.4 mL of TRIzol reagent, vortexed, and incubated for 5 minutes. Each fraction then received 480 µl of chloroform (Thermo Fisher Scientific), incubated for 3 minutes, and centrifuged for 15 min at 20,000 g (4°C). After centrifugation, the clear aqueous phase was placed into a sterile 15 mL conical tube and combined with 1.2 mL of 100% isopropyl alcohol, followed by overnight incubation at -20°C. Samples were then centrifuged at 20,000 g for 30 min (4°C), pellets washed twice with 75% ethanol, and resuspended in DNase/RNase free ddH2O. RNA quality was assessed with the Agilent 4200 TapeStation before being sent to SeqCoast Genomics for sequencing using the Illumina NextSeq 1000 platform. Ribosome depletion was performed using custom rRNA probes designed specifically for ribosomal RNA from *P. longispinus*, including the host and its symbionts, as well as *Sodalis praecaptivus* HS.

Mass spectrometry proteomics of mealybugs

P. longispinus mealybug bacteriomes were dissected into phosphate-buffered saline and immediately flash-frozen in liquid nitrogen. Samples were processed at the Translational Genomics Research Institute (TGen) via biopulverization, followed by lysis in a Precellys homogenizer with soft tissue beads. Samples (approximately 50 µg of protein) were then subjected to in-solution trypsin digestion. Digested peptides were purified using solid-phase extraction on C18 columns dried in a SpeedVac. Peptides were then resuspended in 2% acetonitrile, 0.1% formic acid solution. For peptide fractionation, a kit-based high-pH reversed-phase method was used, yielding eight fractions. Peptides were separated over a 120-minute gradient on a 25 cm Aurora Ultimate C18 column (IonOpticks) using a Vanquish Neo UHPLC system coupled to a Thermo Scientific Orbitrap Eclipse mass spectrometer. Data-dependent acquisition was performed with scan range set to 375 - 1500 m/z. MS1 scans were acquired in the Orbitrap at a resolution of 120,000 FWHM (full width at half-maximum) and MS2 scans were collected in the ion trap.

Bioinformatics

Pseudogene annotation: Pseudogenes were annotated with Pseudofinder (Syberg-Olsen et al., 2022), using, as the reference database, NCBI's non-redundant protein database. In cases where >65% of the gene is retained on the genome, but is broken up into multiple ORFs, that gets labeled as a near-complete pseudogene; we consider these near-complete pseudogenes to be the youngest, most recently formed. If 65% or less of a gene is present on the genome, determined in comparison to the average length of top 50 orthologs from RefSeq, that gene gets labeled as a truncated pseudogene; we consider these pseudogenes to be generally older than near-complete pseudogenes. Finally, regions considered to be intergenic noise by the original gene-calling software used here (Prodigal [Hyatt, 2004], via Prokka [Seeman, 2014]) were aligned against the RefSeq database, and regions with at least 25 significant matches within that database are labeled as cryptic pseudogenes. We consider these pseudogenes to be older than truncated pseudogenes.

Transcriptomics and ribosomal profiling: RNA reads were quality-trimmed using Trimmomatic (Bolger et al., 2014) and aligned against *Sodalis praecaptivus* and endosymbiont genomes using Bowtie2 (Langmead and Salzberg 2012). Per-gene counts data were extracted using HTSeq-Count (Anders et al., 2015) and count values were converted to normalized expression levels (correcting for differences in gene lengths, sequencing effort per sample, and symbiont abundance), expressed as transcripts per million (TPM, Zhao et al., 2021).

For estimation of reads from tmRNA in ribosomal profiling datasets, we searched for similar datasets in NCBI's Sequence Read Archive (SRA). We identified 40 BioProject studies with a total of 438 SRA experiments corresponding to RNA sequencing of ribosome and polysome isolations. Of these, two studies generated data using methods similar to ours (colored green in **Supplemental Table 1**). We used the SRA Toolkit to download the datasets, and tmRNA reads were classified by alignment, using Bowtie2 (Langmead and Salzberg, 2012), against an expanded database of tmRNA sequences from Bacteria and Archaea (Nawrocki et al., 2025). This was repeated for our own ribosomal profiling dataset, where we found that 35% of the *Symbiopectobacterium*-classified reads mapped to the tmRNA database. This is only marginally lower than our estimates based on alignment of reads against the annotated *Symbiopectobacterium* tmRNA sequence.

Statistical comparisons and data visualization: We compared expression levels between pseudogenes and intact genes using Welch's t-test (Welch, 1947), which allows for comparisons between groups of varying sizes (there are more intact genes than pseudogenes) and variances, though there does not appear to be much difference in expression level variation between intact and broken genes. This test was carried out in RStudio, and we used a combination of base R and ggplot2 for visualization.

LC-MS/MS Data Processing: Raw data were processed with Proteome Discoverer (Thermo Fisher Scientific, v3.1). Spectra were searched with the Sequest HT search engine using the species-specific whole proteome FASTA database (enzymatic cleavage set to trypsin, allowing up to two missed cleavages). Precursor mass tolerance was set to 10 ppm and fragment mass tolerance was set to 0.6 Da. Peptide validation was carried out with Percolator (Spivak, 2009), which uses a machine-learning model trained on true (target) and false (decoy) identifications. We set the false discovery rate (FDR) to 1% (Elias and Gygi, 2007). For quantification, precursor-based feature detection and mapping were performed, and ion intensities were extracted for peptide-level abundance quantification. To detect products of ribosome rescue, we generated a custom protein database that was comprised of all possible combinations of peptide + tag from the tmRNA. While this resulted in a large protein database search space for Proteome Discoverer (since we included the tmRNA tag at all possible positions where a protein may stall), we maintained the 1% FDR cutoff to mitigate false positives.

Data Availability

RNA sequencing data generated in this study have been made available via NCBI's Sequence Read Archive (BioProject PRJNA1417914 for *Sodalis praecaptivus* RNA sequencing data and TBD for *Pseudococcus longispinus* RNA data). Mass spectrometry data have been made available via the PRIDE Archive (PXD074041).

Acknowledgements

We would like to thank Tim Karr of the ASU Mass Spectrometry Facility for helpful feedback on the analysis of mass spectrometry data and Rachel Green, Scott Miller, and Tim Wheeler for advice on design of the ribosome profiling experiment. This work was supported by the Howard Hughes Medical Institute and NASA's Interdisciplinary Consortia for Astrobiology Research (grant number 80NSSC23K1357).

REFERENCES

Amiel, E., Lovewell, R. R., O'Toole, G. A., Hogan, D. A., & Berwin, B. (2010). Pseudomonas aeruginosa evasion of phagocytosis is mediated by loss of swimming motility and is independent of flagellum expression. *Infection and Immunity*, *78*(7), 2937–2945. https://doi.org/10.1128/IAI.00144-10

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* , *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Belinky, F., Rogozin, I. B., & Koonin, E. V. (2017). Selection on start codons in prokaryotes and potential compensatory nucleotide substitutions. *Scientific Reports*, *7*(1), 12422. https://doi.org/10.1038/s41598-017-12619-6

Benjak, A., Honap, T. P., Avanzi, C., Becerril-Villanueva, E., Estrada-García, I., Rojas-Espinosa, O., Stone, A. C., & Cole, S. T. (2017). Insights from the Genome Sequence of Mycobacterium lepraemurium: Massive Gene Decay and Reductive Evolution. *mBio*, *8*(5). https://doi.org/10.1128/mBio.01283-17

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* , *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boyd, B. M., James, I., Johnson, K. P., Weiss, R. B., Bush, S. E., Clayton, D. H., & Dale, C. (2024). Stochasticity, determinism, and contingency shape genome evolution of endosymbiotic bacteria. *Nature Communications*, *15*(1), 4571. https://doi.org/10.1038/s41467-024-48784-2

Burke, G. R., & Moran, N. A. (2011). Massive genomic decay in Serratia symbiotica, a recently evolved symbiont of aphids. *Genome Biology and Evolution*, *3*, 195–208. https://doi.org/10.1093/gbe/evr002

Chari, A., Oakeson, K. F., Enomoto, S., Jackson, D. G., Fisher, M. A., & Dale, C. (2015). Phenotypic characterization of Sodalis praecaptivus sp. nov., a close non-insect-associated member of the Sodalis-allied lineage of insect endosymbionts. *International Journal of Systematic and Evolutionary Microbiology*, *65*(Pt 5), 1400–1405. https://doi.org/10.1099/ijs.0.000091

Clayton, A. L., Oakeson, K. F., Gutin, M., Pontes, A., Dunn, D. M., von Niederhausern, A. C., Weiss, R. B., Fisher, M., & Dale, C. (2012). A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genetics*, *8*(11), e1002990. https://doi.org/10.1371/journal.pgen.1002990

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honoré, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., … Barrell, B. G. (2001). Massive gene decay in the leprosy bacillus. *Nature*, *409*(6823), 1007–1011. https://doi.org/10.1038/35059006

Danneels, B., Pinto-Carbó, M., & Carlier, A. (2018). Patterns of nucleotide deletion and insertion inferred from bacterial pseudogenes. *Genome Biology and Evolution*, *10*(7), 1792–1802. https://doi.org/10.1093/gbe/evy140

Engelhardt, F., Tomasch, J., & Häussler, S. (2020). Organism-specific depletion of highly abundant RNA species from bacterial total RNA. *Access Microbiology*, *2*(10), acmi000159. https://doi.org/10.1099/acmi.0.000159

Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, (2007)

Eren, A. M., & Banfield, J. F. (2024). Modern microbiology: Embracing complexity through integration across scales. *Cell*, *187*(19), 5151–5170. https://doi.org/10.1016/j.cell.2024.08.028

Felden, B., Himeno, H., Muto, A., McCutcheon, J. P., Atkins, J. F., & Gesteland, R. F. (1997). Probing the structure of the Escherichia coli 10Sa RNA (tmRNA). *RNA (New York, N.Y.)*, *3*(1), 89–103. https://pubmed.ncbi.nlm.nih.gov/8990402/

Feng, Y., Wang, Z., Chien, K.-Y., Chen, H.-L., Liang, Y.-H., Hua, X., & Chiu, C.-H. (2022). “Pseudo-pseudogenes” in bacterial genomes: Proteogenomics reveals a wide but low protein expression of pseudogenes in Salmonella enterica. *Nucleic Acids Research*, *50*(9), 5158–5170. https://doi.org/10.1093/nar/gkac302

Garber, A. I., Kupper, M., Laetsch, D. R., Weldon, S. R., Ladinsky, M. S., Bjorkman, P. J., & McCutcheon, J. P. (2021). The evolution of interdependence in a four-way mealybug symbiosis. *Genome Biology and Evolution*. https://doi.org/10.1093/gbe/evab123

Goodhead, I., Blow, F., Brownridge, P., Hughes, M., Kenny, J., Krishna, R., McLean, L., Pongchaikul, P., Beynon, R., & Darby, A. C. (2020). Large-scale and significant expression from pseudogenes in Sodalis glossinidius - a facultative bacterial endosymbiont. *Microbial Genomics*, *6*(1). https://doi.org/10.1099/mgen.0.000285

Husnik, F., & McCutcheon, J. P. (2016). Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(37), E5416-24. https://doi.org/10.1073/pnas.1603910113

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*, 119. https://doi.org/10.1186/1471-2105-11-119

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, *324*(5924), 218–223. https://doi.org/10.1126/science.1168978

Janssen, B. D., & Hayes, C. S. (2012). The tmRNA ribosome-rescue system. *Advances in Protein Chemistry and Structural Biology*, *86*, 151–191. https://doi.org/10.1016/B978-0-12-386497-0.00005-0

Keiler, K. C., Waller, P. R., & Sauer, R. T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science (New York, N.Y.)*, *271*(5251), 990–993. https://doi.org/10.1126/science.271.5251.990

Keiler, K. C. (2008). Biology of trans-translation. *Annual Review of Microbiology*, *62*(1), 133–151. https://doi.org/10.1146/annurev.micro.62.081307.162948

Keiler, K. C. (2015). Mechanisms of ribosome rescue in bacteria. *Nature Reviews. Microbiology*, *13*(5), 285–297. https://doi.org/10.1038/nrmicro3438

Kim, W., Hwang, S., Lee, N., Lee, Y., Cho, S., Palsson, B., & Cho, B.-K. (2020). Transcriptome and translatome profiles of *Streptomyces* species in different growth phases. *Scientific Data*, *7*(1), 138. https://doi.org/10.1038/s41597-020-0476-9

Kirchberger, P. C., Schmidt, M. L., & Ochman, H. (2020). The Ingenuity of Bacterial Genomes. *Annual Review of Microbiology*, *74*, 815–834. https://doi.org/10.1146/annurev-micro-020518-115822

Koga, R., & Moran, N. A. (2014). Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *The ISME Journal*, *8*(6), 1237–1246. https://doi.org/10.1038/ismej.2013.235

Kuo, C.-H., & Ochman, H. (2009). Deletional bias across the three domains of life. *Genome Biology and Evolution*, *1*, 145–152. https://doi.org/10.1093/gbe/evp016

Kuo, C.-H., & Ochman, H. (2010). The extinction dynamics of bacterial pseudogenes. *PLoS Genetics*, *6*(8). https://doi.org/10.1371/journal.pgen.1001050

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lerat, E., & Ochman, H. (2005). Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Research*, *33*(10), 3125–3132. https://doi.org/10.1093/nar/gki631

Liu, Y., Harrison, P. M., Kunin, V., & Gerstein, M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biology*, *5*(9), R64. https://doi.org/10.1186/gb-2004-5-9-r64

Liu, Y., Yang, Q., & Zhao, F. (2021). Synonymous but not silent: The Codon usage code for gene expression and protein folding. *Annual Review of Biochemistry*, *90*(1), 375–401. https://doi.org/10.1146/annurev-biochem-071320-112701

Lynch, M., & Marinov, G. K. (2015). The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(51), 15690–15695. https://doi.org/10.1073/pnas.1514974112

Lynch, M., Koskella, B., & Schaack, S. (2006). Mutation pressure and the evolution of organelle genomic architecture. *Science*, *311*(5768), 1727–1730. https://doi.org/10.1126/science.1118884

Mangano, K., Marks, J., Klepacki, D., Saha, C. K., Atkinson, G. C., Vázquez-Laslop, N., & Mankin, A. S. (2022). Context-based sensing of orthosomycin antibiotics by the translating ribosome. *Nature Chemical Biology*, *18*(11), 1277–1286. https://doi.org/10.1038/s41589-022-01138-9

McCutcheon, J. P., & Moran, N. A. (2011). Extreme genome reduction in symbiotic bacteria. *Nature Reviews. Microbiology*, *10*(1), 13–26. https://doi.org/10.1038/nrmicro2670

Mira, A., & Pushker, R. (2005). The silencing of pseudogenes. *Molecular Biology and Evolution*, *22*(11), 2135–2138. https://doi.org/10.1093/molbev/msi209

Mira, A., Ochman, H., & Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends in Genetics: TIG*, *17*(10), 589–596. https://doi.org/10.1016/s0168-9525(01)02447-7

Mohammad, F., Green, R., & Buskirk, A. R. (2019). A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *eLife*, *8*. https://doi.org/10.7554/eLife.42591

Moore, S. D., & Sauer, R. T. (2005). Ribosome rescue: tmRNA tagging activity and capacity in Escherichia coli. *Molecular Microbiology*, *58*(2), 456–466. https://doi.org/10.1111/j.1365-2958.2005.04832.x

Moran, N. A., McLaughlin, H. J., & Sorek, R. (2009). The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*, *323*(5912), 379–382. https://doi.org/10.1126/science.1167140

Nawrocki, E. P., Petrov, A. I., & Williams, K. P. (2025). Expansion of the tmRNA sequence database and new tools for search and visualization. *NAR Genomics and Bioinformatics*, *7*(1), lqaf019. https://doi.org/10.1093/nargab/lqaf019

Nechitaylo, T. Y., Sandoval-Calderón, M., Engl, T., Wielsch, N., Dunn, D. M., Goesmann, A., Strohm, E., Svatoš, A., Dale, C., Weiss, R. B., & Kaltenpoth, M. (2021). Incipient genome erosion and metabolic streamlining for antibiotic production in a defensive symbiont. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(17). https://doi.org/10.1073/pnas.2023047118

Oakeson, K. F., Gil, R., Clayton, A. L., Dunn, D. M., von Niederhausern, A. C., Hamil, C., Aoyagi, A., Duval, B., Baca, A., Silva, F. J., Vallier, A., Jackson, D. G., Latorre, A., Weiss, R. B., Heddi, A., Moya, A., & Dale, C. (2014). Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biology and Evolution*, *6*(1), 76–93. https://doi.org/10.1093/gbe/evt210

Palomba, A., Abbondio, M., Fiorito, G., Uzzau, S., Pagnozzi, D., & Tanca, A. (2021). Comparative evaluation of MaxQuant and Proteome Discoverer MS1-based protein quantification tools. *Journal of Proteome Research*, *20*(7), 3497–3507. https://doi.org/10.1021/acs.jproteome.1c00143

Plague, G. R., Dunbar, H. E., Tran, P. L., & Moran, N. A. (2008). Extensive proliferation of transposable elements in heritable bacterial symbionts. *Journal of Bacteriology*, *190*(2), 777–779. https://doi.org/10.1128/JB.01082-07

Posit team (2025). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/

Ray, B. K., & Apirion, D. (1979). Characterization of 10S RNA: a new stable rna molecule from Escherichia coli. *Molecular & General Genetics: MGG*, *174*(1), 25–32. https://doi.org/10.1007/bf00433301

Roberts, M., & Josephs, E. B. (2023). Weaker selection on genes with treatment-specific expression consistent with a limit on plasticity evolution in Arabidopsis thaliana. *Genetics*, *224*(2), iyad074. https://doi.org/10.1093/genetics/iyad074

Rocha, E. P. C. (2004). Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Research*, *14*(11), 2279–2286. https://doi.org/10.1101/gr.2896904

Roche, E. D., & Sauer, R. T. (1999). SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. *The EMBO Journal*, *18*(16), 4579–4589. https://doi.org/10.1093/emboj/18.16.4579

Saito, K., Kratzat, H., Campbell, A., Buschauer, R., Burroughs, A. M., Berninghausen, O., Aravind, L., Green, R., Beckmann, R., & Buskirk, A. R. (2022). Ribosome collisions induce mRNA cleavage and ribosome rescue in bacteria. *Nature*, *603*(7901), 503–508. https://doi.org/10.1038/s41586-022-04416-7

Samatova, E., Daberger, J., Liutkute, M., & Rodnina, M. V. (2020). Translational control by ribosome pausing in bacteria: How a non-uniform pace of translation affects protein production and folding. *Frontiers in Microbiology*, *11*, 619430. https://doi.org/10.3389/fmicb.2020.619430

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* , *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Smith, C., Canestrari, J. G., Wang, A. J., Champion, M. M., Derbyshire, K. M., Gray, T. A., & Wade, J. T. (2022). Pervasive translation in Mycobacterium tuberculosis. *eLife*, *11*. https://doi.org/10.7554/eLife.73980

Suzuki, K., Nakata, N., Bang, P. D., Ishii, N., & Makino, M. (2006). High-level expression of pseudogenes in Mycobacterium leprae. *FEMS Microbiology Letters*, *259*(2), 208–214. https://doi.org/10.1111/j.1574-6968.2006.00276.x

Syberg-Olsen, M. J., Garber, A. I., Keeling, P. J., McCutcheon, J. P., & Husnik, F. (2022). Pseudofinder: Detection of pseudogenes in prokaryotic genomes. *Molecular Biology and Evolution*, *39*(7), msac153. https://doi.org/10.1093/molbev/msac153

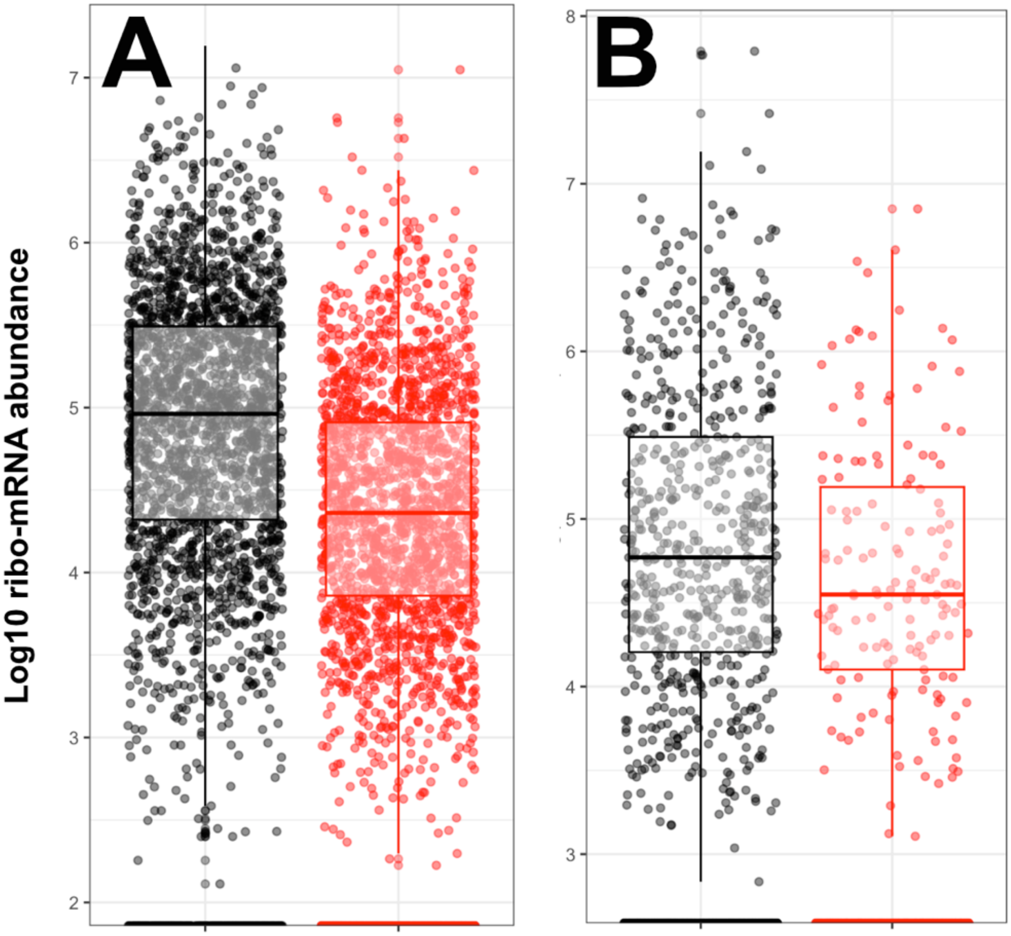
Toh, H., Weiss, B. L., Perkin, S. A. H., Yamashita, A., Oshima, K., Hattori, M., & Aksoy, S. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host. *Genome Research*, *16*(2), 149–156. https://doi.org/10.1101/gr.4106106

Williams, D. L., Slayden, R. A., Amin, A., Martinez, A. N., Pittman, T. L., Mira, A., Mitra, A., Nagaraja, V., Morrison, N. E., Moraes, M., & Gillis, T. P. (2009). Implications of high level pseudogene transcription in Mycobacterium leprae. *BMC Genomics*, *10*(1), 397. https://doi.org/10.1186/1471-2164-10-397

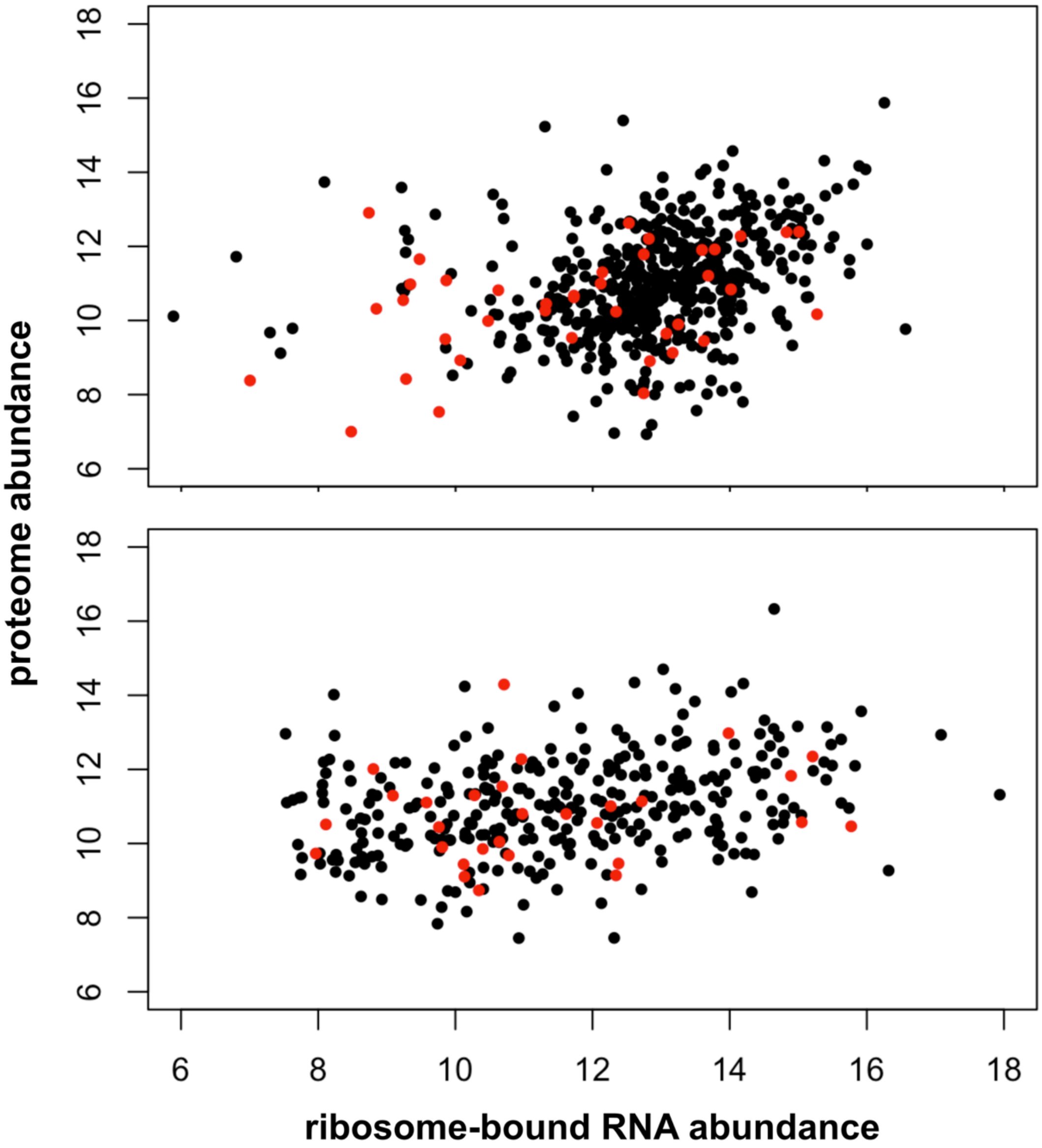
Yannai, A., Katz, S., & Hershberg, R. (2018). The Codon usage of lowly expressed genes is subject to natural selection. *Genome Biology and Evolution*, *10*(5), 1237–1246. https://doi.org/10.1093/gbe/evy084

Zhao, Y., Li, M.-C., Konaté, M. M., Chen, L., Das, B., Karlovich, C., Williams, P. M., Evrard, Y. A., Doroshow, J. H., & McShane, L. M. (2021). TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. *Journal of Translational Medicine*, *19*(1), 269. https://doi.org/10.1186/s12967-021-02936-w

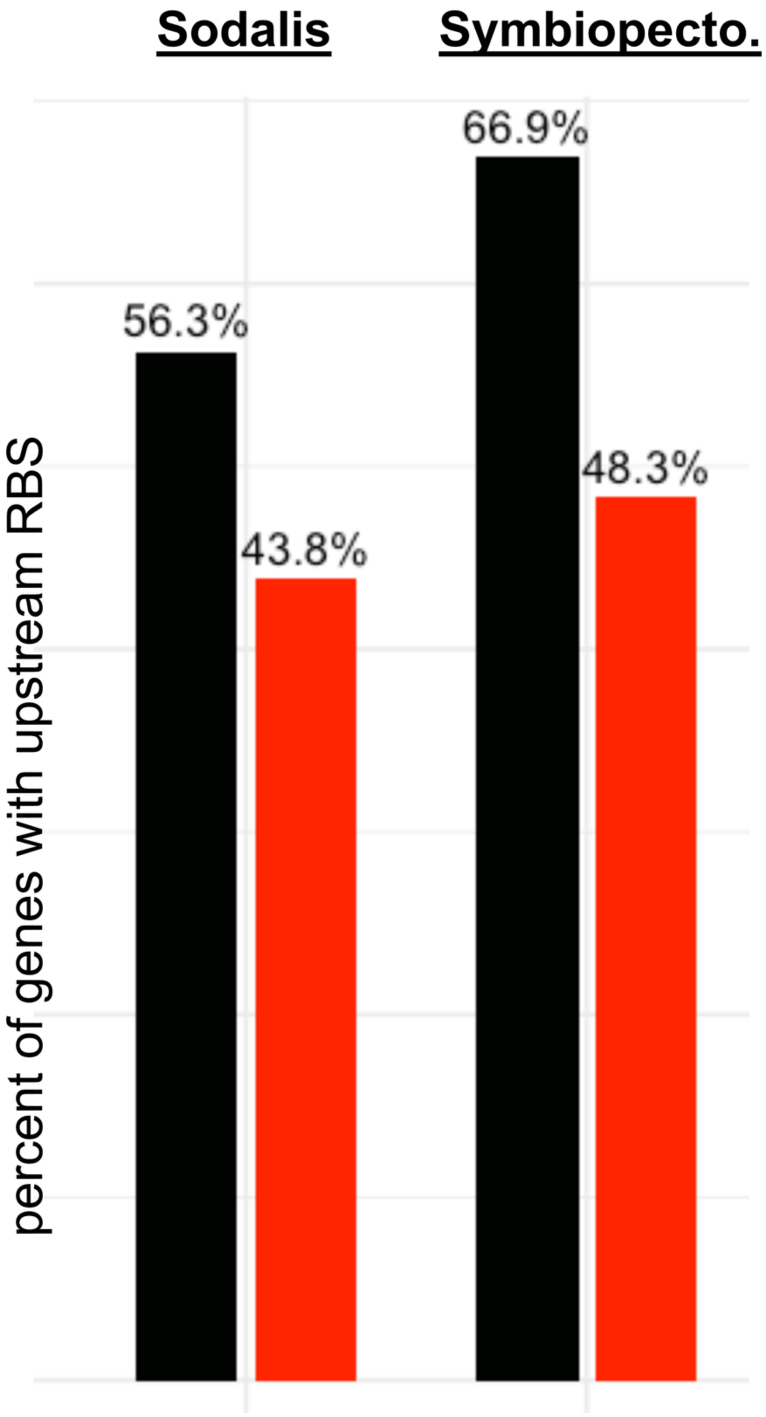
SUPPLEMENTAL FIGURES



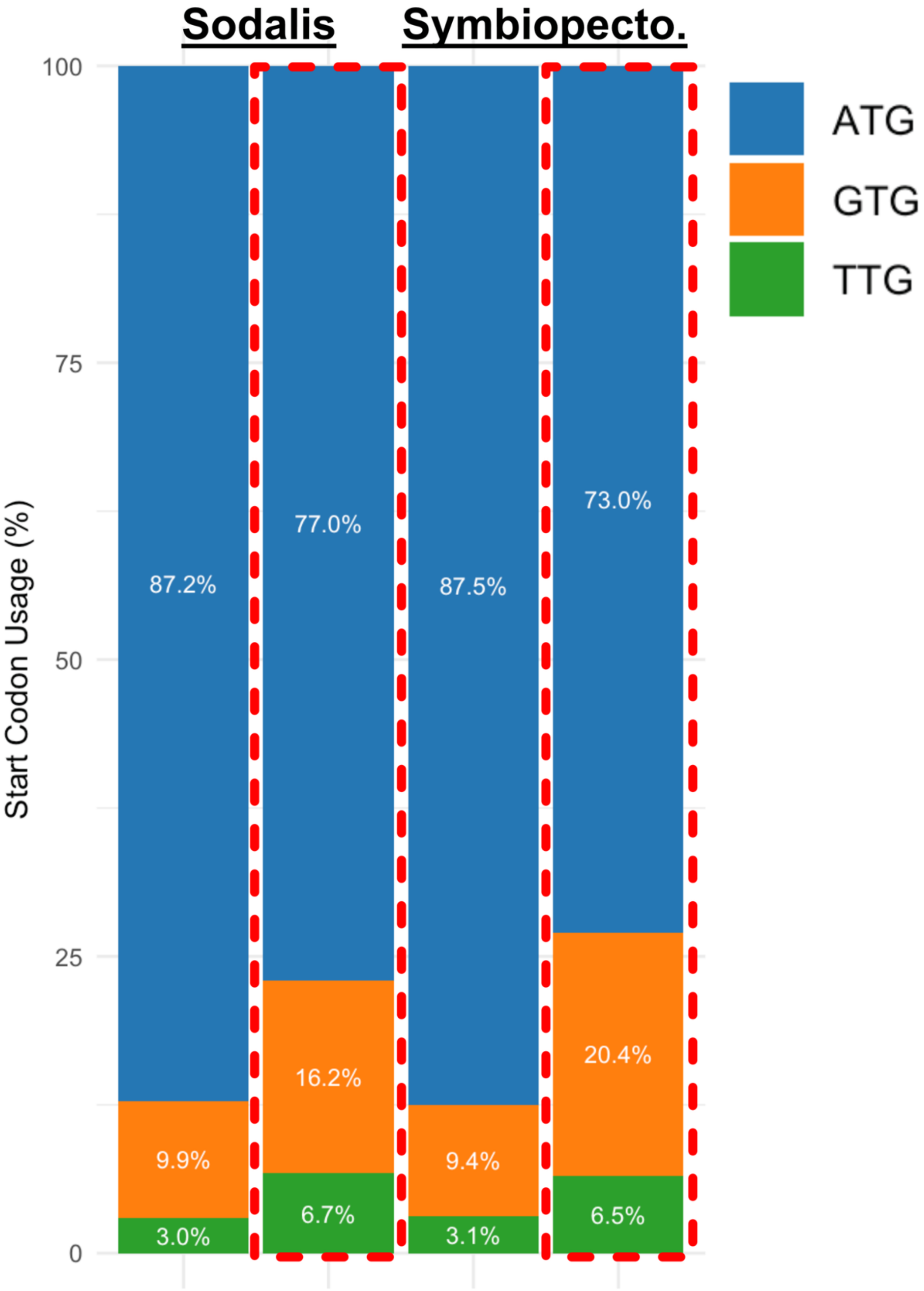
Supplemental Figure 1: Percentage of intact (black) and pseudogenes (red) with upstream ribosomal binding sites, as predicted via Prodigal.



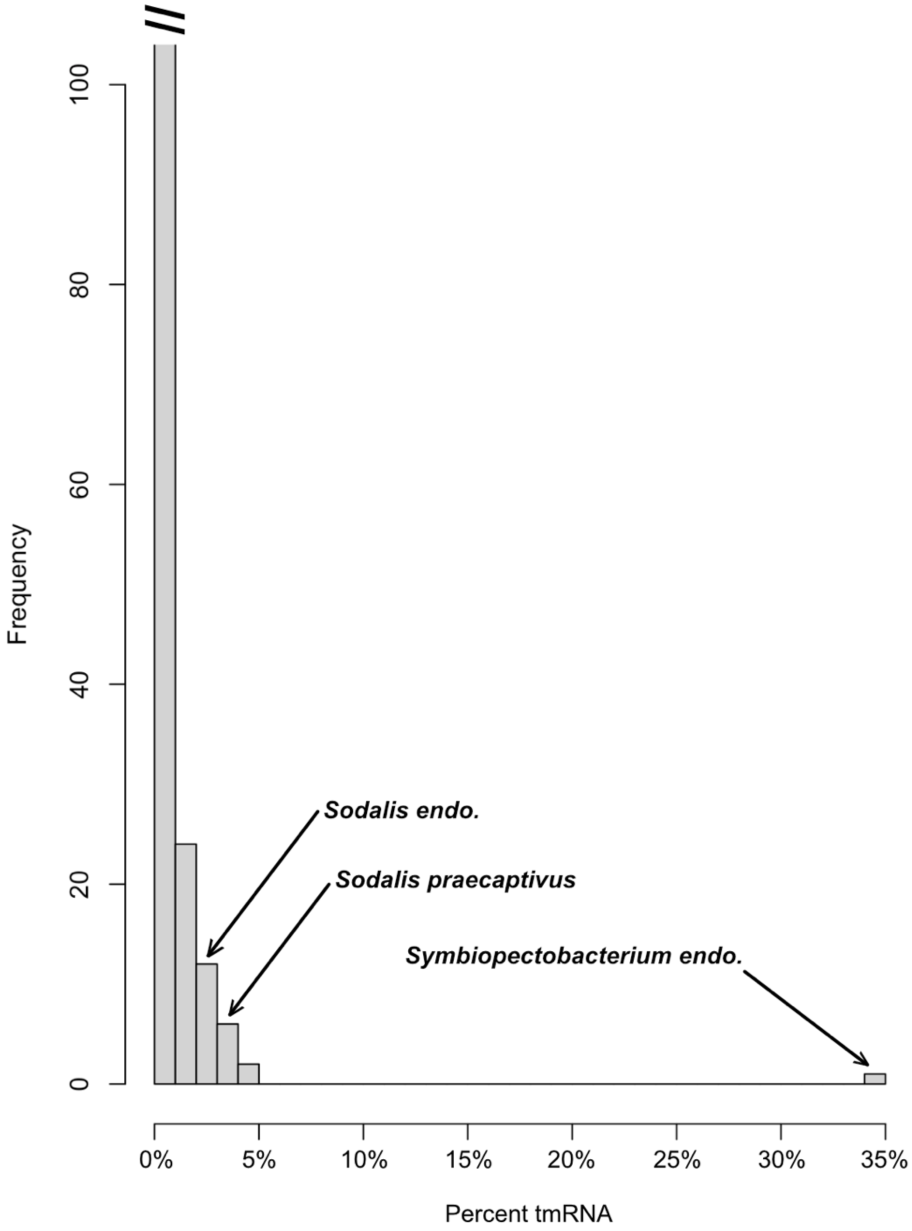
Supplemental Figure 2: Comparison of ribosome-bound RNA levels with protein abundance in *Symbiopectobacterium endo*. (top) and *Sodalis endo*. (bottom). Abundances are shown on a normalized log2 scale, with pseudogenes colored red.



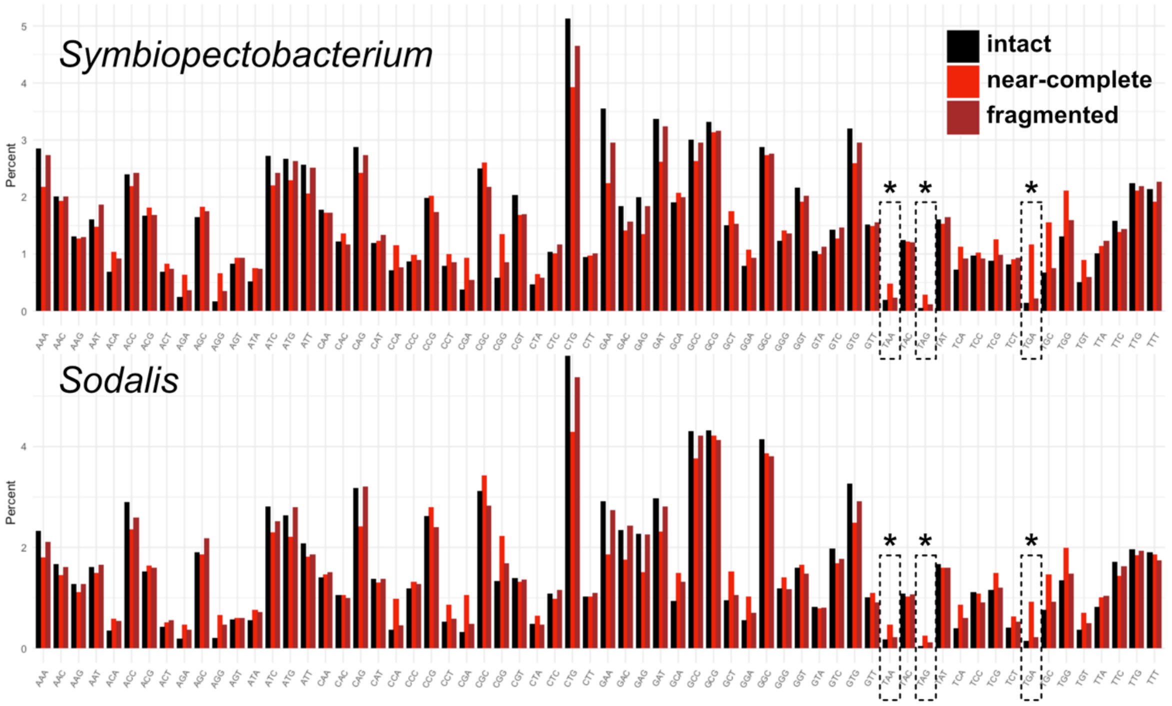
Supplemental Figure 3: Percentage of intact (black) and pseudogenes (red) with upstream ribosomal binding sites, as predicted via Prodigal.



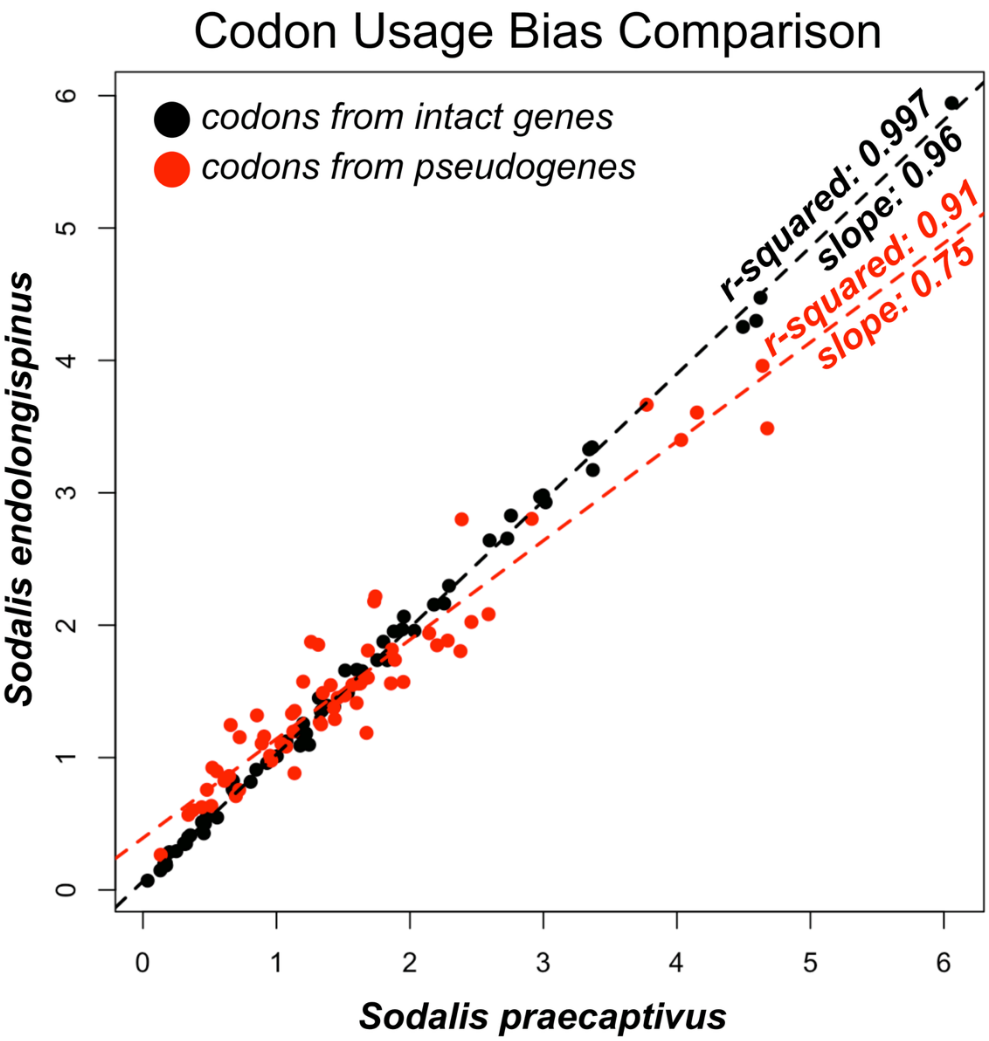
Supplemental Figure 4: Percent of intact and pseudogenes (enclosed in dotted red boxes) that start with each of three types of start codons.



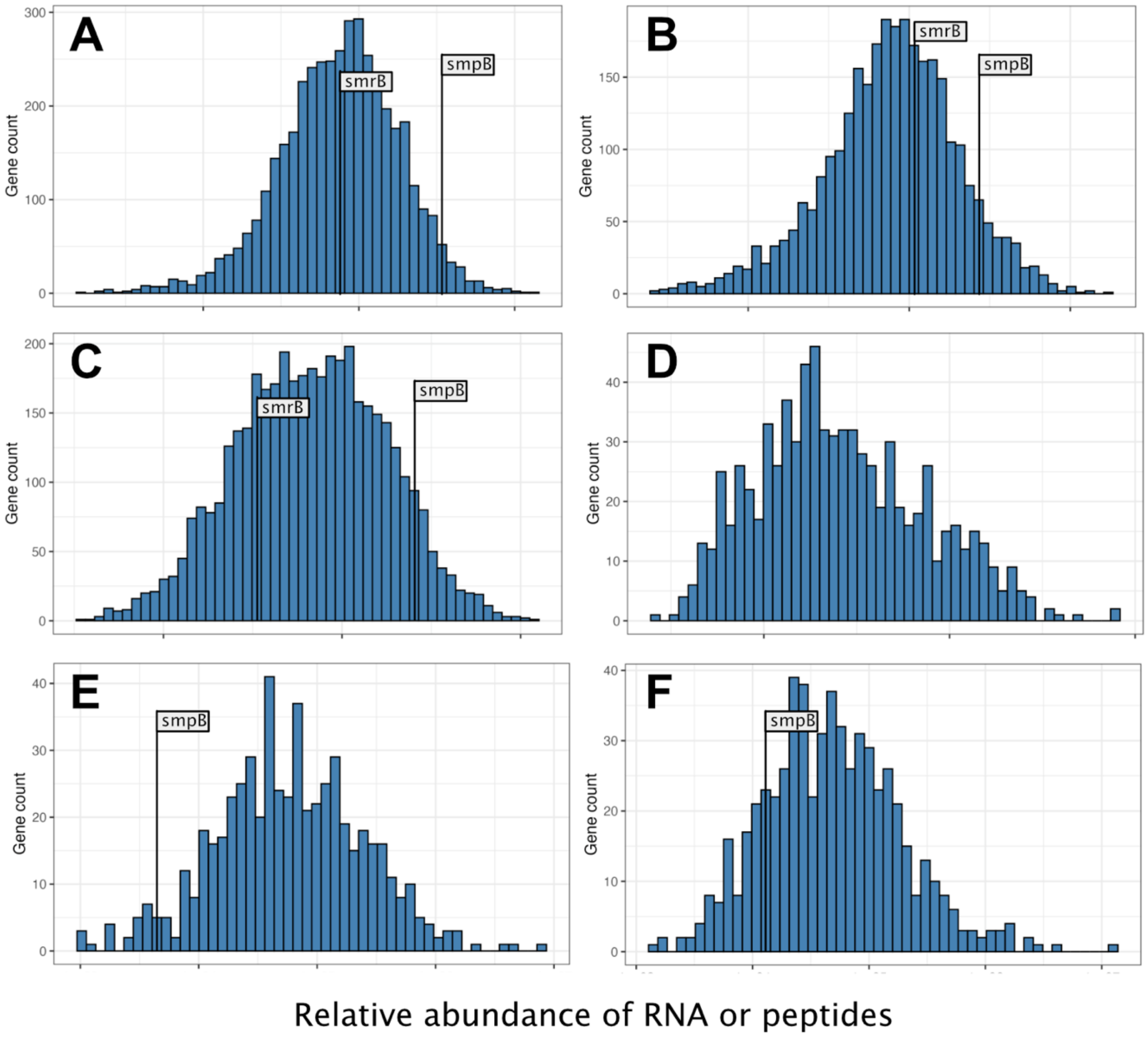
Supplemental Figure 5: Histogram showing the percent of reads classified as tmRNA from 438 ribosomal profiling datasets.



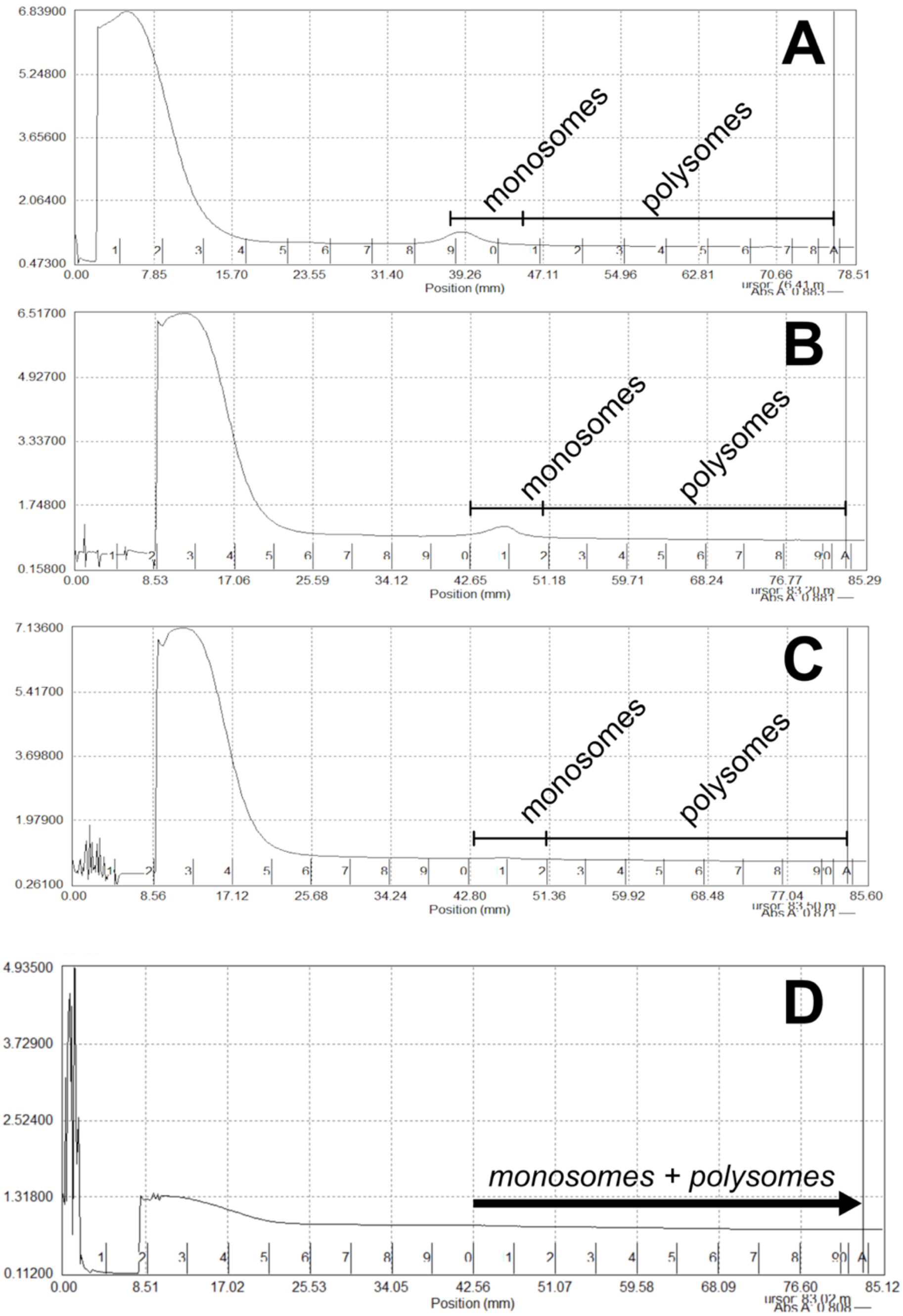
Supplemental Figure 6: Codon usage percentages across all intact genes (black), in comparison with two types of pseudogenes: near-complete and truncated. Y-axis shows percent across all codons (within-category). Highlighted with asterisks are stop codons TAG, TTG, and TGA.



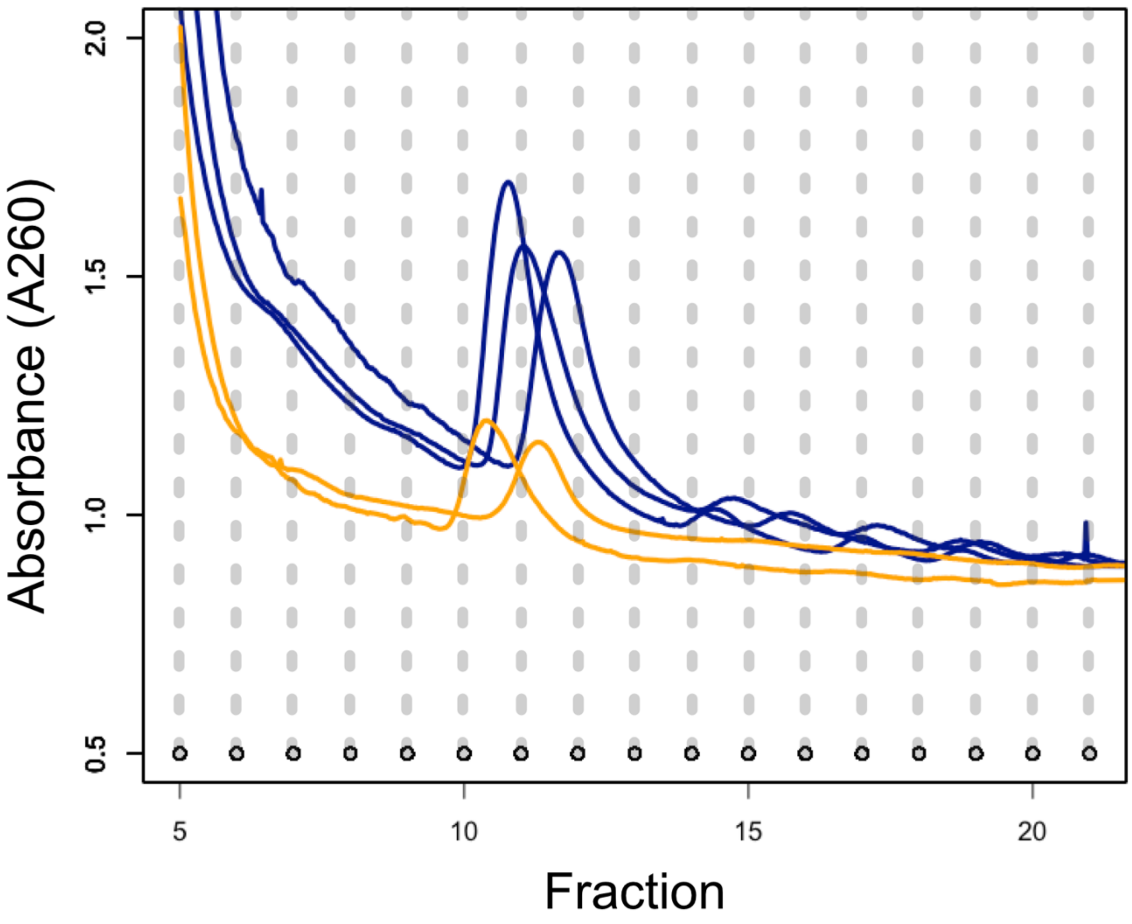
Supplemental Figure 7: Comparison of codon usages between *Sodalis endo*. and its closest free-living relative *Sodalis praecaptivus*. Codons from pseudogenes in *Sodalis endo*. are shown in red. Linear model results superimposed over each regression.



Supplemental Figure 8: Histograms showing the distribution of RNA or peptide levels in *Symbiopectobacterium endo*. and *Sodalis endo*. Top row shows relative levels among whole transcriptomes in A) *Symbiopectobacterium* and B) *Sodalis*. Middle row shows relative levels among ribosome-copurified RNA in C) *Symbiopectobacterium* and D) *Sodalis*. Bottom row shows relative protein abundance levels in E) *Symbiopectobacterium* and F) *Sodalis*.



Supplemental Figure 9: Absorbance measurements (A260), shown on the y-axis, taken during fractionation of whole-insect *Pseudococcus longispinus* samples (A-C) and P. longispinus bacteriomes (D). Fraction numbers are shown on the x-axis, with the top of the centrifuge tube corresponding to fraction 1 and the bottom (including pellet) corresponding to fraction 20. Monosome peaks are visible in panels A and B starting at fraction 10, which is the highest fraction we used for downstream sequencing and analysis (i.e., we used fractions 10-20 to infer ribosome-bound RNAs).



Supplemental Figure 10: Absorbance measurements taken during fractionation of a 10-50% sucrose gradient after ultracentrifugation. Only fractions 5-20 are shown. *Sodalis praecaptivus* HS samples are shown in blue (triplicate), and two whole-insect *P. longispinus* samples (from panels A and B in Supplemental Figure 7) are shown in orange. Each dot on the bottom corresponds to an individual fraction from the sucrose gradient, collected automatically via the gradient fractionator.