Evidence of multifaceted functions of codon usage in translation within the model beetle *Tribolium castaneum*

Authors: Carrie A. Whittle[1], Arpita Kulkarni[1], Cassandra G. Extavour[1,2]

[1] Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge MA 02138, USA Phone: 617-495-1564

[2]Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge MA 02138, USA

Corresponding Author: Cassandra Extavour, Email: extavour@oeb.harvard.edu

Running title: Codon use shapes translation in beetles

Abstract

Synonymous codon use is non-random. Codons most used in highly transcribed genes, often called optimal codons, typically have high gene counts of matching tRNA genes (tRNA abundance) and promote accurate and/or efficient translation. Non-optimal codons, those least used in highly expressed genes, may also affect translation. In multicellular organisms, codon optimality may vary among tissues. At present however, codon use remains poorly understood in multicellular organisms. Here, we studied codon usage of genes highly transcribed in germ line (testis, ovary) and somatic tissues (gonadectomized males and females) of the beetle *Tribolium castaneum*. The results demonstrate that: 1) the majority of optimal codons were organism-wide, the same in all tissues, and had numerous matching tRNA gene copies ($Opt_{\uparrow tRNA}$), consistent with translational selection; 2) some optimal codons varied among tissues, suggesting tissue-specific tRNA populations; 3) wobble tRNA were required for translation of certain optimal codons ($Opt\text{-}codon_{wobble}$), possibly allowing precise translation and/or protein folding; and 4) remarkably, some non-optimal codons had abundant tRNA genes ($Nonopt\text{-}codon_{\uparrow tRNAs}$), and genes using those codons were tightly linked to ribosomal and stress-response functions. Thus, $Nonopt\text{-}codon_{\uparrow tRNAs}$ codons may regulate translation of specific genes. Together, the evidence suggests that codon use and tRNA genes regulate multiple translational processes in beetles.

Keywords: Optimal codons, non-optimal codons, translational selection, translation regulation, tRNA genes

1. Introduction

In protein coding genes, the synonymous codons of amino acids are not used randomly. Biases in codon usage are thought to result from selection for translational efficiency and/or accuracy.[1-9] Mutational pressures can also shape codon usage.[5,10-13] Translational selection in many organisms has been supported by findings that the highly transcribed genes preferentially use a subset of codons, often described as "optimal" codons,[2,6,12-18] and has been observed in bacteria,[5,6,17] fungi,[16,19,20] plants[2,14,21] and animals, including spiders[22] and insects (e.g., *Drosophila, Aedes, Anopheles*, *Gryllus*, *Oncopeltus,* and weakly observed in *Bombyx* [2,15,23-27]). Whole-genome data show that optimal codons typically have correspondingly high numbers of iso-accepting tRNA gene copies in the genome, reflecting an organism's relative tRNA abundance,[1,5,6,19,20,28] and is consistent with selection for translational optimization.[1,4,5,18,20,29-33] The utility of tRNA gene number to quantify organismal tRNA abundance has been supported *in vivo* in bacteria and eukaryotes.[28,34,35] For instance, the addition of tRNA genes for a codon of a specific amino acid to the *E. coli* genome markedly improved translation rates of genes containing that amino acid.[28] In this regard, the increased use of optimal codons in highly transcribed genes,[2,5,14] and the correspondence of these codons to abundant tRNA genes,[1,4] suggest that selection may favor translational optimization.

In contrast to unicellular systems, multicellular organisms measuring codon usage can be complicated by the plurality of tissues, as optimal codons and tRNA populations may vary among tissue types.[36-38] For instance, cellular tRNA abundances can vary among tissues or cell types for at least some codons,[37,39,40] suggesting that translational selection may differ among tissues.[37] This has also been supported by findings of some variation in codon use of genes transcribed in different tissues in the few organisms studied to date. For example, in the plant *Arabidopsis* use of specific codons in a gene depends on the tissue type in which it is maximally expressed, suggesting this species has localized tRNA populations,[38] a pattern that has also been proposed for rice.[41] Although similar studies in metazoans have been rare, a recent investigation in *D. melanogaster* showed that codons associated with elevated expression were not universal across tissues. For example, AAT was more commonly used than AAC for Asn in some tissues (e.g., testis, hindgut), while TGT was favored over TGC for Cys in the salivary glands, that was suggested to provide evidence of tissue-specific tRNA populations.[36] Additional studies are warranted to determine the universality of distinct optimal codon identities in various tissues of an organism. In particular, the germ line and somatic tissues comprise contrasts of significant interest, as the former directly determines an organism's reproductive success and fitness and

experiences haploid selection in the meiotic and sex cells, such that translational optimization may be particularly relevant to those tissues.

While much attention has been focused on optimal codons in the literature, growing experimental research, largely from single-celled models or in vitro systems, suggests that non-optimal codons, those codons least used in highly transcribed genes (and/or codons defined as "rare" in some studies), can also play significant regulatory roles in translation.[34,42] In yeast for example, it was shown that cells altered their tRNA populations under stress and had increased levels of tRNAs that matched the rare codons found in stress-response genes, thus allowing the preferential translation of those mRNAs under stressful conditions, without any change in mRNA abundance.[43] Findings in cyanobacteria have indicated that circadian rhythms are regulated post-transcriptionally based on non-optimized codon use in genes of the *kaiABC1* cluster.[44] Further, non-optimal codons have been shown to slow rates of translational elongation and to control ribosome traffic on mRNA, which allows proper co-translational protein folding and/or functionality, based on *in vitro* cell-free translation systems from *Neurospora*[7] and *Drosophila*.[9] Non-optimal codons have also been found to facilitate co-translational protein folding in various yeast models.[45] These data show that the use of even one or few rare codon(s) in a gene may markedly affect its translation, depending on the tRNA pool, suggesting that the supply-demand relationship between non-optimal codons and their matching tRNA abundances could comprise an adaptive mechanism of translational regulation.[34,43-47] To further understand this phenomenon, genomics and molecular evolution research on codon usage patterns in animal systems should expand beyond the typical focus on optimal codons, and specifically include assessments of non-optimal codons, and their relationships to tRNA genes.

In addition to non-optimal codons *per se*, some studies have indicated that the use of codons that have no matching tRNA, and obligately require wobble codon-anticodon tRNAs (wobbly at the third nucleotide of the codon) may also influence translation.[34] For instance, an investigation in four divergent eukaryotes found that the relative translation levels of cell-cycling gene mRNAs during various stages of the cell cycle depended on the frequency of codons that had no corresponding tRNA gene copies in the genome and thus required wobble tRNA.[48] Further, experimental research in yeast, human cells, and nematodes has shown that obligatory use of wobble tRNA decelerates translational elongation by slowing ribosomal translocation on the mRNA.[34,49,50] In this regard, the use of codons that require wobble tRNA could have a significant effect on translational dynamics, particularly in slowing translation,[34] and thus should also be considered in studies of codon usage patterns in an organism.

A metazoan species providing a promising pathway for the comprehensive study of codon usage in a multicellular system is the Coleopteran rust red flour beetle *Tribolium castaneum*. *T. castaneum* is a long standing model for genetics and developmental biology, has a well characterized genome,[18,51,52] and is estimated to have diverged from the fellow insect *Drosophila* approximately 300 Mya.[53-57] While a prior pioneering study had identified a putative list of optimal codons for *T. castaneum,*[18] the approach used in that study involved correlation analyses between codon frequency and expression level. Given that this method has been thought to often be poorly suited to revealing optimal codons, defined as those most common in highly transcribed genes,[1,5,58] analyses of codon use in this taxon would benefit from being revisited with alternative methods. Optimal codons can be most readily revealed via direct contrasts of codon usage in the highest versus lowest expressed genes in the genome, also known as the contrast method.[13-17,21,24,58] At present, like most multicellular model organisms, a multifaceted integrative approach has not yet been applied to assessments of codon usage in this beetle taxon, including the identification of optimal and non-optimal codons in highly transcribed genes at an organism-wide level, and within the somatic versus germ line tissues, nor have assessments been available of the links been such codon usage and tRNA gene counts, wobble tRNA, and gene functionality.

In the present study, we address these outstanding issues on codon usage in *T. castaneum* using genome-wide protein-sequence datasets (CDS) and large-scale transcriptome datasets from the male and female germ lines and somatic tissues (testes, ovaries, gonadectomized (GT-) males and GT-females).[59] From these data, we rigorously study optimal and non-optimal codons in this taxon, and their relationships to tRNA abundances and gene ontology. From these analyses, we report strong evidence for organism-wide optimal codons in all four tissue types and both sexes. The majority of these optimal codons have abundant matching tRNAs ($Opt_{\uparrow tRNA}$), consistent with pervasive translational selection for efficient and/or accurate protein synthesis in this species. A minority of optimal codons vary among the four tissues, suggesting small, but potentially meaningful, differences in tRNA populations between tissue types. Crucially, we report that a subset of the optimal codons did not have direct tRNA matches and obligately required wobble tRNA for translation ($Opt\text{-}codon_{wobble}$), which we propose may comprise a mechanism for slowing translation for accuracy or protein-folding purposes. Finally, we find that a number of non-optimal codons unexpectedly have abundant perfectly matching tRNA gene copies ($Nonopt\text{-}codon_{\uparrow tRNAs}$) and that these rare codons are preferentially used in genes with specific functions, including ribosomal protein genes and stress response genes. Thus, we hypothesize that the use of

129 codons with Nonopt-codon$_{\uparrow tRNAs}$ status may be a potential mechanism to ensure preferential translation
130 of specific gene mRNAs. Collectively, our results reveal the multiple roles of codon usage in this beetle,
131 suggesting not just pervasive selection for the use of specific codons in highly transcribed genes for
132 efficient and/or accurate translation, but also translational regulatory roles of wobble codons and of non-
133 optimal codons.
134
135 **2. Materials and Methods**
136
137 **2.1. *T. castaneum* CDS**
138 The annotated CDS of our main target species *T. castaneum* (v.5.2) were downloaded from
139 Ensembl Metazoa (http://metazoa.ensembl.org) and are also available at BeetleBase[51,52]). The full CDS
140 per gene (longest CDS per gene) was used for the study of codon usage. The full genome and its
141 descriptive GFF file was also downloaded for assessments.
142
143 **2.2. Biological samples and RNA-seq**
144 We aimed to determine the expression level (FPKM) for each of 16,434 genes in *T. castaneum*
145 for germ line and somatic tissues. For this we used the large-scale RNA-seq datasets for the ovaries,
146 testes, GT-females and GT-males shown in Supplementary Table S1.[59] The *T. castaneum* specimens
147 were provided by the Brown lab at KSU (https://www.k-state.edu/biology/people/tenure/brown/).
148 Samples were grown under standard laboratory conditions until adulthood and tissue dissections were
149 then performed on unmated adults (a total of 150 animals per sex per biological replicate), and RNA was
150 extracted and processed for RNA-seq, as described previously.[59]
151
152 **2.3. Gene expression**
153 The RNA-seq reads (76bp) per sample were trimmed of adapters and poor-quality bases using
154 the BBduk available from the Joint Genome Initiative (https://jgi.doe.gov/data-and-tools/bbtools/) set at
155 default parameters.
156 Gene expression level was determined for the 16,434 genes (CDS) as FPKM after mapping each
157 RNA-seq dataset per tissue to the full CDS list for each species using Geneious Read Mapper[60], which
158 yielded highly similar results as other common mappers such as BBmap (https://jgi.doe.gov/data-and-
159 tools/bbtools/). The average FPKM across samples per tissue type (Supplementary Table S1) was used

to measure expression per tissue. FPKM values were highly correlated between replicates of each sample type (Spearman's Ranked R>0.9, P<2X10^{-7}).

2.4. Identification of optimal and non-optimal Codons

For identification of the optimal codons, we measured the relative synonymous codon usage (RSCU) per codon per amino acid for each gene under study using CAICal.[61] RSCU values indicate the relative usage of a codon in a synonymous codon family, and values >1 indicate favored usage and <1 indicate unfavored usage as compared other codons. For each of the 18 amino acids in the genetic code with synonymous codons (note that Trp and Met only have one codon each), we identified the optimal codon using the contrast method.[13-15,17,21,24,58,62] For this, we determined the difference in RSCU (ΔRSCU) per codon between genes with the highest 5% versus the lowest 5% expression. The primary optimal codon for each amino acid was defined as the codon with the highest and statistically significant positive ΔRSCU value, indicating preferred usage in highly transcribed genes.[13-15,17,21,24,58,62] The primary non-optimal codon per amino acid was defined as the codon with the largest negative and statistically significant ΔRSCU value, indicating low usage in highly transcribed genes. Statistical significance per codon was applied using a t-test between RSCU values across all genes for high versus low expressed genes.

As the literature reflects some variation in terminology among studies to date, we explicitly define the term "optimal codons" herein as those codons most used in highly transcribed genes based on ΔRSCU, which infers an innate advantage of the codon under high transcription. Then, we secondarily assessed each optimal codon's correspondence to the number of matching (codon-anticodon) tRNA genes in order to test their role in translational accuracy/efficiency[1,4,5,18,20,29-33] or possible other functions (e.g., wobble codons for translational slowing). For non-optimal codons a similar approach was used wherein the non-optimal codons were identified based solely on ΔRSCU, and their relationships to tRNA were then assessed.

The frequency of optimal codons (Fop) is a measure of the degree of optimal codon usage per gene.[6] Fop was determined in CodonW[63] using the primary optimal codons identified herein. Fop was also determined using the primary optimal codons previously identified by Williford and Demuth 2012.[18] As multiple codons per amino acid were classified as optimal in that assessment, we defined each primary optimal codon from their study as that with the strongest average positive correlation across tissues for measuring Fop.

For an additional layer stringency, we wished to exclude the possibility that expression-mediated mutational-biases towards specific nucleotides, which have been observed to some extent in certain organisms to date (e.g., *E. coli*, humans[64,65]), contribute towards codon differences among high and low expressed genes herein. For this, we extracted all introns for every gene in the genome (those with introns) using the GFF file available (see section 2.1). Introns are thought to be mostly selectively neutral,[18,66] and thus the nucleotide content should reflect any underlying mutational pressures in the genome, and on the nucleotide composition of synonymous codons in an organism.[13,18,66] If mutational pressures on introns are not associated with gene expression level, it will exclude this factor in shaping optimal codons, and further affirm the role of selection. All introns that were >50bp were extracted as the region between exons and were concatenated per gene. The association between GC content and expression level were assessed using a scatter plot and Spearman's ranked R.

2.5. Identification of tRNA Genes

To assess whether or how the optimal and non-optimal codons were related to the tRNA gene copy number, we determined the number of iso-accepting tRNA genes per codon in the genome (*T. castaneum* v. 5.2) using tRNA-scan SE.[18,52,67] The list of tRNA gene numbers identified in the current genome version was identical to that reported previously[18] and is shown in Table 1.

2.6. GO Functions

The predicted GO functions were determined using Panther[68] using the option for *T. castaneum* as species.

2.7. Data Availability

The CDS and genome v. 5.2 for *T. castaneum* are available at Ensembl Metazoa (http://metazoa.ensembl.org). RNA-seq data for all samples from *T. castaneum* described in Supplementary Table S1 are available at the SRA database under Bio-project number PRJNA564136.

3. Results and Discussion

3.1. Optimal codons in *T. castaneum*

We first report the organism-wide, or global, optimal codon per amino acid for *T. castaneum* using ΔRSCU and the average expression levels of all annotated genes across all four studied tissue

types (testis, ovary, GT-male, GT-female) in Table 1. The primary optimal codon was defined as the codon with the largest positive ΔRSCU between highly and lowly transcribed genes and with $P<0.05$, and was found for 17 of the 18 amino acids with synonymous codons. Seven primary optimal codons ended in T, three in A, five in C and two in G. We noted that Ile had two codons with nearly identical ΔRSCU values. Further, CAC for His showed signs of optimal codon usage in several individual tissues (see following section), and including this codon yields a study-wide total of 18 optimal codons (Table 1). Thus, the patterns in Table 1 indicate that selection pressures have favored the use of a specific subset of codons in highly expressed genes[5] (for results on non-optimal codons see section 3.4 below).

While the striking use of specific optimal codons in genes under high expression levels in Table 1 in itself provides evidence of selection on codon usage, we wished to include additional layers of stringency to affirm the role of selection in favoring these codons. First, we determined the frequency of optimal codons (Fop), a measure of the degree of optimal codon usage per gene,[6] for all studied genes in the genome (N=16,434). As shown in Fig. 1A, we found that the Fop increased from genes with low (top 5% in the genome), to moderate (5 to 95%), to high (top 5%) expression levels (Ranked ANOVA and Dunn's paired test $P<0.05$). As low and high expressed genes were used to identify the optimal codons, the Fop was expectedly lowest and highest in those categories of genes respectively. Importantly however, moderately expressed genes, which were not used to identify the optimal codons, showed intermediate Fop values, suggesting a gradation in optimal codon use with greater expression, and thus infers a genome-wide selective pattern. Second, as codon usage can vary with protein length in some eukaryotes,[2,5,69] we repeated the assessment in Fig. 1A using genes with similar CDS lengths, which we binned into short (<150 codons), medium (≥150, <300), and long CDS (≥300). For each of these three length categories, we found the same stepwise increase of Fop values with expression level (Ranked-ANOVAs $P<0.001$). Thus, the link between expression and optimal codons cannot be explained by protein length. Third, from examination of introns, wherein nucleotide content is mostly shaped by mutational pressures,[18,66,70] we found that the GC (and thus AT) content of introns was fully uncorrelated to gene expression level (Spearman's correlation $R= -0.09$, Fig. 1B),[71,72] and thus indicates an absence of expression-mediated mutational biases[12,64,65,70] in this species. Further to this point, unlike some organisms wherein optimal codons typically end in only two or three types of nucleotides,[2,14,21,24] all four nucleotides are represented at the terminal position of optimal codons of this species (Table 1); this also excludes mutational biases in shaping the optimal codons this taxon.[5] Taken together, these

supplemental analyses add additional rigor and allow us to conclude that the optimal codons in Table 1 reflect a history of selection pressures favoring their usage in highly expressed genes.

3.2. Most, but not all, optimal codons are the same across germ line and somatic tissues

In order to compare optimal codon usage among the tissues under study, we next determined the optimal codons (using ΔRSCU) using genes with high versus low expression (top and lowest 5%) separately for each of the four individual tissue types, ovaries, testes, GT-females and GT-males. For rigor in this assessment, we identified the subset of genes in the top 5% expression class that were only in the top category for one tissue type (and were not in the top 5% expression in any of the other three tissues), to discern whether or not there was a tissue effect on optimal codons. Under these criteria, we identified 372, 450, 444, and 272 genes for analysis, for ovaries, testis, GT-females and GT-males respectively. This allowed us to specifically assess the codon usage of genes that were maximally transcribed only in one individual tissue, as it has been found that if tissue-type has an effect on codon use, this effect is most apt to be evident in its highly transcribed genes[38]. The results for ΔRSCU per tissue type are shown in Table 1. We report that 15 of the 18 primary optimal codons (including His) from the organism-wide assessment were identified as having the same optimal codon in three, or all four, of the individual tissue-types (Table 1). Thus, the vast majority of primary optimal codons were the same in these divergent tissues, including male and female germ lines and somatic tissue types.

However, several significant differences were also observed among tissues. For example, a male-specific primary optimal codon was identified for the amino acid Phe (with two synonymous codons), as the codon TTC was optimal in the testes and GT-males, but not in the ovaries or GT-females (Table 1). Similarly, a GT-male-specific primary optimal codon ATC was identified for Ile (with three synonymous codons), where ATT was optimal for the other three tissues. In turn, an ovary-specific optimal codon was evident for Pro (with four synonymous codons), as the primary optimal codon was CCC in all tissues except for the ovaries, where it was CCT. In addition, a GT-female optimal codon was identified for Lys (two synonymous codons), where AAG was optimal in the ovaries, testes, and GT-males, but its alternate codon AAA was optimal for GT-females. These examples show that the primary optimal codon varies among tissue type in this beetle, and thus this pattern suggests that translational selection regimes, and thus corresponding tRNA populations, may also vary among tissues.[36]

These present results are consistent with the few available studies of tissue-specific codon usages and translational selection from the fellow insect *D. melanogaster*[36] and in studied plants[38,41] (note that although some evidence suggests humans have tissue-specific optimal codons, this has been debated, and may largely be an effect of the GC content of isochores, which exist in those organisms[73,74]). Together, while the vast majority of optimal codons are shared across tissues in these beetles, non-negligible differences are observed between tissues and sexes. Direct quantification of tRNAs in cells or tissues has been mostly restricted to date to lab models of bacteria, yeast or *in vitro* human cell lines,[37,39,40,43,75] and the accuracy and limitations of the various approaches (based on microarrays, Northern blot, quantitative PCR, RNA-seq) remains debated[40,43,76,77]. Nevertheless, the development of robust methods to sequence tRNAs that are applicable to non-traditional model organisms will allow further tests of whether or how tRNA expression levels vary with tissues in *T. castaneum*, as is strongly suggested by these results.[36]

3.3. A majority of organism-wide optimal codons have high tRNA gene copy numbers

Given the minimal differences among tissues, for our remaining analyses we focus on the organism-wide optimal codon usages (Table 1). The number of tRNA gene copies in the genome has commonly been used as a measure of the relative abundance of each tRNA species.[1,4,18,20,29,30,48] If optimal codon usage were consistently a result of selection in response to abundant tRNAs, then the primary optimal codon per amino acid should also have high relative tRNA gene frequency. When using the organism-wide optimal codon list (Table 1), we found that 12 of the primary optimal codons also had the highest, or near the highest tRNA gene counts of all codons per amino acid, GCT (Ala), AGA (Arg), AAC (Asn), CAA (Gln), GAA (Glu), ATT (Ile), TTG (Leu), AAG (Lys), TTC (Phe), ACT (Thr), TAC (Tyr), and GTT (Val). Further, while the positive ΔRSCU of CAC for His was not statistically significant using the organism-wide assessment ($P=0.26$), this codon was optimal when individually considered in the ovaries, GT-females and GT-males ($P<0.05$), and had seven matching tRNA genes. Thus, when including CAC for His as a codon with optimal status, yields a study-wide total of 13 of the 18 primary optimal codons that have plentiful matching tRNA genes. In other words, a majority of optimal codons have $Opt_{\uparrow tRNA}$ status. These results suggest translational selection for accuracy and/or efficiency[1,4] across a majority of amino acids in this beetle.

312 ***Hypothesis 1: Optimal codons use wobble tRNA to resolve conflict of high translation with sequence***
313 ***fidelity***

314 While 13 optimal codons had a high number of direct tRNA matches as expected under selection
315 for optimization of efficient and accurate translation, for the remaining five amino acids, a much
316 different pattern was observed. Specifically, the primary optimal codon (highly used in abundant
317 transcripts) had no direct matching tRNA-genes, and a wobble tRNA (shown in Table 1) must thus be
318 employed for translation of these codons (denoted as Opt-codon$_{wobble}$). For instance, Opt-codon$_{wobble}$
319 status was observed for the amino acids Asp (GAT), Cys (TGT), Gly (GGT), Pro (CCC) and Ser (AGT).
320 Thus, this result shows that while these identified optimal codons are preferred in highly transcribed
321 genes, their innate benefit cannot be due to having abundant direct matching tRNA, and thus another
322 mechanism must explain their high usage. Further, as shown in Supplementary Text File 1 and Fig.
323 S1, within the group of highly transcribed genes, each of these five codons with Opt-codon$_{wobble}$ status
324 showed strong associations with protein length, inferring putatively significant roles in the translation of
325 mRNAs encoding longer proteins.

326 Experimental studies in bacteria and eukaryotic models have shown that codons using wobble
327 tRNA act to slow translation by decelerating the translocation of ribosomes on mRNA.[34,49,50] In addition,
328 a study of the genomes of various eukaryotes (humans, yeast, *Arabidopsis*) have indicated that cell-cycle
329 genes had high usage of codons that had no matching tRNA genes in the genome, and thus must employ
330 wobble tRNA, which inherently have lower codon-anticodon binding affinity than those codons with
331 perfect matches.[48] The differential use of codons using wobble tRNA in cell-cycle genes, combined with
332 potential oscillations in tRNA abundances, were proposed to differentially regulate the translation rates
333 of gene mRNAs during various stages of the cell cycle.[48] Further, this was speculated to possibly
334 comprise a broader evolutionarily conserved phenomenon for translational regulation in eukaryotes.[48] In
335 addition, the usage of wobble-tRNAs in a gene could have some parallel functions to the use of non-
336 optimal codons (those non-optimal codons with low tRNA abundance) which can prevent jamming of
337 multiple ribosomes during the initiation of translation,[35] and/or slow or pause translation during
338 elongation, which would facilitate accurate protein-folding.[7,9,39,78] In this regard, the results from these
339 various studies suggest that the slowing of translation that is induced by wobble-tRNA[34,49,50] could
340 comprise an evolutionarily conserved mechanism shaping various aspects of translation.

341 Taken together, we hypothesize here that for this beetle, the use of codons with Opt-codon$_{wobble}$
342 status in highly expressed genes comprises a mechanism to slow or pause translation at various sites,

which may lead to increased accuracy of translation or allow co-translational protein folding,[49] factors particularly significant for longer proteins. In addition, the high frequency of codons with Opt-codon$_{wobble}$ status in genes with abundant mRNAs (Table 1), suggests that these codons might also play a significant role in post-transcriptional differential regulation of protein levels[48] in these beetles. Additional studies of protein levels of genes with high usage of codons with Opt-codon$_{wobble}$ status will be needed to further test this aspect of the hypothesis.

3.4. Certain non-optimal codons have abundant tRNA genes

Herein, we defined the primary non-optimal codon per amino acid stringently as the codon with the largest negative ΔRSCU per amino acid, rather than simply all codons that were not optimal. Using these data, we assessed whether those codons with low usage in highly transcribed genes also exhibit few tRNA gene copies, as might be expected if codon usage is mostly shaped by translational selection for efficient and accurate translation (i.e., for adaptation of optimal codons and tRNA abundance). The organism-wide primary non-optimal codons (per amino acid) are shown in Table 1.

The results showed that some non-optimal codons, as expected, had low numbers of matching tRNA genes (e.g., two tRNA genes for ACG (Thr), ATA (Ile), and TTA (Leu), one for CCG (Pro), and zero for CGG (Arg)). Unexpectedly, however, certain non-optimal codons had relatively moderate to high tRNA gene abundance (denoted as Nonopt-codon$_{\uparrow tRNAs}$). For instance, for Arg, whilst the codon CGG had no tRNA gene copies, its sister non-optimal codon CGA (ΔRSCU= -0.290 and -0.265 respectively) had four tRNA gene matches. For Gly, both the primary and secondary non-optimal codons GGC and GGA (-0.104 and -0.077 respectively) had eight and 15 matching tRNA gene copies respectively. For Val, the primary non-optimal codon GTA had five tRNA genes, only slightly lower than the seven observed for its optimal codon GTT. We noted, that if we relaxed our definition of a non-optimal codon to consider any codon that is not optimal, we found that some of those codons also had many corresponding tRNA genes. For example, for Pro the non-optimal codon CCA (which had a weak and nonsignificant positive ΔRSCU value, +0.029, and thus would not have satisfied our strict definition of having the largest negative ΔRSCU for this amino acid) had 13 tRNA genes, an extraordinarily high value compared with other codons. Moreover, for Asp, the (less stringently) defined non-optimal codon GAC had ten matching tRNA copies. Collectively, it is evident that codons that are not the optimal codons in this taxon are not inevitably linked to a low abundance of matching tRNA genes, and rather in some cases exhibit high matching tRNA gene counts. Thus, these patterns suggest it

is possible that non-optimal codons with elevated tRNAs play a specific regulatory role for highly transcribed genes.

A recent study in yeast has indicated that stress genes may preferentially use non-optimal codons that have abundant iso-accepting tRNA genes, to increase effective gene expression by promoting their translation over other proteins rather than affecting mRNA levels.[43] Based on this notion, we hypothesize here that codons with Nonopt-codon$_{\uparrow tRNAs}$ status in *T. castaneum* may regulate the translation of abundant mRNAs of proteins with specific functions in this beetle. To further evaluate this possibility, we examined the predicted gene ontology functions of the highly transcribed genes that had relatively elevated usage of non-optimal codons with abundant tRNAs.

Hypothesis 2: Non-optimal codons post-transcriptionally regulate translation based on protein functions

We assessed the GO functions of highly transcribed genes (top 5% in the genome from the organism-wide analyses across all four tissues, N=822; and a cutoff of 103.3 FPKM) that had relatively elevated use of codons with Nonopt-codon$_{\uparrow tRNAs}$ status (Table 1). For this assessment, rather than assess all strictly defined non-optimal codons, we chose as examples the codons GGC for Gly, GTA for Val, and CGA for Arg. These three codons were defined as non-optimal by our strict definition (having a large negative and statistically significant ΔRSCU, Table 1) and had substantial matching tRNA gene copy counts (four to eight tRNA genes each). These codons also had negative ΔRSCU values in all four of the tissue types studied (Table 1), indicating they consistently have non-favored status in this organism.

For the amino acid Gly, we identified those highly transcribed genes that had RSCU values for GGC of >1.5. An RSCU value of one is expected for each of the four Gly codons under equal usage, and thus values of 1.5 to 4 for GGC are relatively high. A total of 20.4% of the highly transcribed gene set was in this class. As shown in Table 2, these genes included those involved in oxidative stress response, such as Peroxiredoxin, and those involved in olfactory activity. Thus, we speculate that these types of genes, which use codons with Nonopt-codon$_{\uparrow tRNAs}$ status, will exhibit less tRNA competition during translation elongation than those genes that use codons with few or no matching tRNA genes, such as the fellow Gly codon GGG (with only one tRNA match), or even those genes using non-optimal codons for other amino acids, such as CCG for Pro (with one tRNA match) (Table 1). In addition, we found that genes with elevated GGC frequency encoded numerous (N=15) ribosomal proteins. Thus, this finding

suggests that usage of the non-optimal codon GGC may shape translation via a second mechanism: namely, by shaping the cellular abundance of specific ribosomal proteins *per se*, which are needed for translation. In this regard, the non-optimal codon usage profiles in Gly appear consistent with a hypothesis wherein the usage of GGC regulates the translation of a subset of genes in this taxon, and may even regulate translation rates *per se* via effects on certain ribosomal proteins.

In terms of Val, those genes with high expression (top 5% in the genome), very rarely used the identified primary non-optimal codon GTA. In fact, only 5.1% of the 822 highly transcribed genes had GTA RSCU values >1.5, an extraordinarily low frequency. Those that did exhibit RSCU values >1.5 included genes involved in cytoskeleton functions and actin synthesis, such as Cofilin/actin-depolymerizing factor homolog-like protein and profilin, as well as a p53-related cell death protein, and a number of uncharacterized proteins (Table 2). For Arg, which has six synonymous codons, genes with RSCU values >1.5 for the non-optimal codon CGA included genes involved in olfactory signaling and with cytoskeleton roles (Table 2). It is particularly noteworthy that unlike the genes with elevated RSCU (e.g. GGC for Gly), which included abundant ribosomal protein genes, no ribosomal protein genes were among those with elevated frequency of GTA in Val or CGA for Arg. Thus, the ribosomal proteins in particular appear to be strongly connected to the usage of the non-optimal GGC Gly codon, and thus we speculate that this codon may be particularly essential to their regulation.

As mentioned above, prior data have suggested that non-optimal codons, when combined with low tRNA abundance, can play important regulatory roles by preventing the jamming of multiple ribosomes during initiation of translation, or slowing translation elongation and facilitating precise protein-folding.[7,9,35,39,78] The present study, however, shows an additional, and much different, plausible effect of non-optimal codons. Specifically, we show that the use of non-optimal codons with abundant tRNA genes (Nonopt-codon$_{\uparrow tRNAs}$) is tightly linked to predicted gene functionality (Table 2), and thus these codons may be likely to contribute to the preferential translation of mRNAs of specific types of genes. This notion agrees with recent experimental data in yeast suggesting that non-optimal or rare codons in stress genes promote their preferential translation in response to stress-induced changes in tRNA pools.[43] Herein, however, given that abundant tRNA gene copies are available in the genome for codons with Nonopt-codon$_{\uparrow tRNAs}$ status (and thus tRNAs should be consistently abundant in cells), we speculate that the use of these non-optimal codons in ribosomal protein and stress genes (Table 2) likely acts as a mechanism to ensure their preferential translation among the various mRNAs within cells at an organism-wide level, independent of environmental or tissue fluctuations in tRNA.

Collectively, our data on codons with Nonopt-codon$_{\uparrow tRNAs}$ status add to the growing support for a mechanism wherein non-optimal or rare codons, combined with elevated tRNA abundances, significantly shape translational regulation in eukaryotes.[34,43,44,46] Further study in beetles, possibly including assessments of protein abundance of genes with elevated usage of codons with Nonopt-codon$_{\uparrow tRNAs}$ status and with high usage of non-optimal codons with rare tRNA genes, will help unravel the relationships between non-optimal codon usage and translation. In addition, *in vivo* quantification of the tRNA populations in diverse tissue types in this beetle,[37,39,40,43,75] will help affirm whether these codons consistently exhibit high tRNA abundances, which could promote their preferential translation at an organism-wide level.

3.5. Comparison of Present Optimal Codon List to a Prior Report

On a final note, it is worthwhile to mention here that the optimal codon list we present in Table 1 differs from that previously reported in *T. castaneum.*[18] The previous report used a correlation method to determine optimal codons, and a comparison of the present primary optimal codon list in Table 1 (for the whole organism analyses) to those earlier findings is shown in Supplementary Table S2. We found that only nine of the 18 primary optimal codons identified herein, were also identified as optimal by the previous study under the correlation method[18], even when we used very loose criteria for defining a match to that prior assessment (that is, considering all optimal codons that were defined at any level under the correlation method, regardless of whether they were the primary, secondary, or tertiary optimal codon[18], as a match to our primary optimal codon). It has been previously argued that the use of a correlation approach can often yield a misleading list of optimal codons.[58] Further, the R values observed for the codons defined as optimal using the prior correlation method were typically <0.1 (the highest value was 0.237, Supplementary Table S2)[18]. A range of such low values, even when statistically significant, is sometimes considered a very weak or absent correlation ($R<0.3$),[71,72] and thus may not be conducive to revealing codons most often used in highly transcribed genes, as was the goal here. Moreover, we found that increased gene expression level (organism-wide expression) was not positively connected to the Fop when using the optimal codons (primary optimal codon defined as strongest correlation) identified under the prior correlation method.[18] Rather, as shown in Supplementary Fig. S2, we found only mild variation in Fop among expression classes, and Fop was reduced in low and high expressed genes as compared to moderately expressed (Ranked ANOVA and Dunn's $P<0.05$), trends inconsistent with a persistent connection between Fop and expression level. However, we did find

a strong connection between expression level and Fop using the optimal codons identified herein (Table 1, Fig. 1A). The method of employing ΔRSCU between high and low expressed genes has repeatedly been shown effective for specifically revealing the optimal codons, defined as those preferentially used in the most highly transcribed genes in the genome,[14,15,17,21,24,58] as was the present objective. Thus, the optimal codons defined herein are those most often used in highly transcribed genes, and were used for all our analyses (Table 1).

3.6 Conclusions

The present study has revealed the complex dynamics of codon usage in the multicellular beetle model system *T. castaneum*. We found that the majority of optimal codons in this animal model are shared at the organism-wide level and match tRNA with abundant gene copies, supporting the presence of species-wide translational selection for efficient and/or accurate translation. However, we also showed that a non-negligible subset of optimal codons varied among the four tissue types, suggesting a likelihood of tissue- and sex-specific tRNA populations, and thus localized translational selection. Based on codon optimality status and tRNA gene copies, we propose two hypotheses. The first hypothesis suggests that the usage of codons with Opt-codon$_{wobble}$ status in highly transcribed genes in this beetle has evolved as a mechanism that slows translation, which could increase precision of translation and/or protein folding. The second hypothesis proposes that usage of codons with Nonopt-codon$_{\uparrow tRNAs}$ status is as a mechanism that promotes high translation of mRNA of genes with specific cellular functions, which we show here to include stress response and ribosomal protein genes.

Further study in beetles, including assessments of cellular protein levels of genes using codons with Opt-codon$_{wobble}$ and Nonopt-codon$_{\uparrow tRNAs}$ status in germ line and somatic tissues, will help further unravel their potential roles in translation regulation. In addition, *in vivo* quantification of the tRNA populations in various tissue types and under stressful conditions in this beetle, as this methodology improves,[37,39,40,43,75,76] will provide additional valuable insights into tRNA population stability and variation between tissues. At present, most non-traditional multicellular organisms have not had as many protocols optimized for lab-based experimental or transgenic research of codon optimization, including rates of translation elongation, protein folding, tRNA-charging, or codon-anticodon tRNA binding, as compared to the established widely studied single-celled models or *in vitro* cell lines.[7,28,34,50] We have shown here, however, that a multifaceted approach using analyses of gene expression, tRNA genes,

497 tissue-type, and gene functionality can be used to suggest how codon usage shapes translational
498 optimization and regulation in metazoans.
499
500 **Acknowledgments**
501 The authors thank Prof. Sue Brown at KSU for generously providing samples of *T. castaneum*
502 for this study. This work was supported by funds from Harvard University. Thanks also to the Extavour
503 lab members for discussions and the Bauer core sequencing facility at Harvard for generating RNA-seq
504 data.

References

1. Duret, L. 2000, tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends in Genetics*, **16**, 287-289.

2. Duret, L. and Mouchiroud, D. 1999, Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 4482-4487.

3. Stoletzki, N. and Eyre-Walker, A. 2007, Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Molecular biology and evolution*, **24**, 374-381.

4. Rocha, E. P. 2004, Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res*, **14**, 2279-2286.

5. Akashi, H. 2001, Gene expression and molecular evolution. *Current Opinion in Genetics & Development*, **11**, 660-666.

6. Ikemura, T. 1981, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol*, **151**, 389-409.

7. Yu, C. H., Dang, Y., Zhou, Z., et al. 2015, Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Mol Cell*, **59**, 744-754.

8. Hershberg, R. and Petrov, D. A. 2008, Selection on codon bias. *Annu Rev Genet*, **42**, 287-299.

9. Zhao, F., Yu, C. H. and Liu, Y. 2017, Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. *Nucleic Acids Res*, **45**, 8484-8492.

10. Sueoka, N. 1988, Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 2653-2657.

11. Sharp, P. M., Averof, M., Lloyd, A. T., Matassi, G. and Peden, J. F. 1995, DNA sequence evolution: the sounds of silence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **349**, 241-247.

12. Comeron, J. M. 2004, Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. *Genetics*, **167**, 1293-1304.

13. Rao, Y., Wu, G., Wang, Z., Chai, X., Nie, Q. and Zhang, X. 2011, Mutation bias is the driving force of codon usage in the Gallus gallus genome. *DNA Res*, **18**, 499-512.

14. Ingvarsson, P. K. 2008, Molecular evolution of synonymous codon usage in Populus. *BMC Evolutionary Biology*, **8**, 307.

15. Cutter, A. D., Wasmuth, J. D. and Blaxter, M. L. 2006, The evolution of biased codon and amino acid usage in nematode genomes. *Molecular biology and evolution*, **23**, 2303-2315.

16. Whittle, C. A., Sun, Y. and Johannesson, H. 2011, Evolution of synonymous codon usage in *Neurospora tetrasperma* and *Neurospora discreta*. *Genome biology and Evolution*, **3**, 332-343.

17. Satapathy, S. S., Powdel, B. R., Buragohain, A. K. and Ray, S. K. 2016, Discrepancy among the synonymous codons with respect to their selection as optimal codon in bacteria. *DNA Res*.

18. Williford, A. and Demuth, J. P. 2012, Gene expression levels are correlated with synonymous codon usage, amino acid composition, and gene architecture in the red flour beetle, Tribolium castaneum. *Mol Biol Evol*, **29**, 3755-3766.

19. Sharp, P. M., Tuohy, T. M. and Mosurski, K. R. 1986, Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic acids research*, **14**, 5125-5143.

20. Percudani, R., Pavesi, A. and Ottonello, S. 1997, Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae. *J Mol Biol*, **268**, 322-330.

21. Qiu, S., Bergero, R., Zeng, K. and Charlesworth, D. 2011, Patterns of codon usage bias in Silene latifolia. *Mol Biol Evol*, **28**, 771-780.

22. Whittle, C. A. and Extavour, C. G. 2016, Expression-Linked Patterns of Codon Usage, Amino Acid Frequency, and Protein Length in the Basally Branching Arthropod Parasteatoda tepidariorum. *Genome Biol Evol*, **8**, 2722-2736.

23. Whittle, C. A. and Extavour, C. G. 2015, Codon and amino acid usage are shaped by selection across divergent model organisms of the Pancrustacea. *G3: Genes, Genomes, Genetics* **5**, 2307-2321.

24. Whittle, C. A. and Extavour, C. G. 2017, Rapid Evolution of Ovarian-Biased Genes in the Yellow Fever Mosquito (Aedes aegypti). *Genetics*, **206**, 2119-2137.

25. Behura, S. K. and Severson, D. W. 2011, Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in Aedes aegypti and Anopheles gambiae. *Insect Mol Biol*, **20**, 177-187.

26. Stenico, M., Lloyd, A. T. and Sharp, P. M. 1994, Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic acids research*, **22**, 2437-2446.

27. Jia, X., Liu, S., Zheng, H., et al. 2015, Non-uniqueness of factors constraint on the codon usage in Bombyx mori. *BMC Genomics*, **16**, 356.

28. Du, M. Z., Wei, W., Qin, L., et al. 2017, Co-adaption of tRNA gene copy number and amino acid usage influences translation rates in three life domains. *DNA Res*, **24**, 623-633.

29. Ikemura, T. 1985, Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, **2**, 13-34.

30. Cognat, V., Deragon, J. M., Vinogradova, E., Salinas, T., Remacle, C. and Marechal-Drouard, L. 2008, On the evolution and expression of Chlamydomonas reinhardtii nucleus-encoded transfer RNA genes. *Genetics*, **179**, 113-123.

31. Moriyama, E. N. and Powell, J. R. 1997, Codon usage bias and tRNA abundance in Drosophila. *J Mol Evol*, **45**, 514-523.

32. Behura, S. K. and Severson, D. W. 2011, Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in *Aedes aegypti* and *Anopheles gambiae*. *Insect Molecular Biology*, **20**, 177-187.

33. Powell, J. R. and Moriyama, E. N. 1997, Evolution of codon usage bias in Drosophila. *Proc Natl Acad Sci U S A*, **94**, 7784-7790.

34. Stein, K. C. and Frydman, J. 2019, The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. *J Biol Chem*, **294**, 2076-2084.

35. Tuller, T., Carmi, A., Vestsigian, K., et al. 2010, An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, **141**, 344-354.

36. Payne, B. L. and Alvarez-Ponce, D. 2019, Codon usage differences among genes expressed in different tissues of Drosophila melanogaster. *Genome Biol Evol*.

37. Dittmar, K. A., Goodenbour, J. M. and Pan, T. 2006, Tissue-specific differences in human transfer RNA expression. *PLoS Genet*, **2**, e221.

38. Camiolo, S., Farina, L. and Porceddu, A. 2012, The relation of codon bias to tissue-specific gene expression in Arabidopsis thaliana. *Genetics*, **192**, 641-649.

39. Quax, T., Claassens, N., Soll D and van der Ooost, J. 2015, Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell*, **59**, 149-161.

40. Goodarzi, H., Nguyen, H. C. B., Zhang, S., Dill, B. D., Molina, H. and Tavazoie, S. F. 2016, Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell*, **165**, 1416-1427.

41. Liu, Q. 2012, Mutational bias and translational selection shaping the codon usage pattern of tissue-specific genes in rice. *PLoS One*, **7**, e48295.

42. Novoa, E. M. and Ribas de Pouplana, L. 2012, Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet*, **28**, 574-581.

43. Torrent, M., Chalancon, G., de Groot, N. S., Wuster, A. and Madan Babu, M. 2018, Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Sci Signal*, **11**.

44. Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y. and Johnson, C. H. 2013, Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature*, **495**, 116-120.

45. Pechmann, S. and Frydman, J. 2013, Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*, **20**, 237-243.

46. Gingold, H., Dahan, O. and Pilpel, Y. 2012, Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. *Nucleic Acids Res*, **40**, 10053-10063.

47. Presnyak, V., Alhusaini, N., Chen, Y. H., et al. 2015, Codon optimality is a major determinant of mRNA stability. *Cell*, **160**, 1111-1124.

48. Frenkel-Morgenstern, M., Danon, T., Christian, T., et al. 2012, Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. *Mol Syst Biol*, **8**, 572.

49. Stadler, M. and Fire, A. 2011, Wobble base-pairing slows in vivo translation elongation in metazoans. *RNA*, **17**, 2063-2073.

50. Letzring, D. P., Dean, K. M. and Grayhack, E. J. 2010, Control of translation efficiency in yeast by codon-anticodon interactions. *RNA*, **16**, 2516-2528.

51. Wang, L., Wang, S., Li, Y., Paradesi, M. S. and Brown, S. J. 2007, BeetleBase: the model organism database for Tribolium castaneum. *Nucleic Acids Res*, **35**, D476-479.

52. Tribolium Genome Sequencing, C., Richards, S., Gibbs, R. A., et al. 2008, The genome of the model beetle and pest Tribolium castaneum. *Nature*, **452**, 949-955.

53. Brown, S. J., Shippy, T. D., Miller, S., et al. 2009, The red flour beetle, Tribolium castaneum (Coleoptera): a model for studies of development and pest biology. *Cold Spring Harb Protoc*, **2009**, pdb emo126.

54. Savard, J., Marques-Souza, H., Aranda, M. and Tautz, D. 2006, A segmentation gene in tribolium produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell*, **126**, 559-569.

55. Denell, R. 2008, Establishment of tribolium as a genetic model system and its early contributions to evo-devo. *Genetics*, **180**, 1779-1786.

56. Choe, C. P., Stellabotte, F. and Brown, S. J. 2017, Regulation and function of odd-paired in Tribolium segmentation. *Dev Genes Evol*, **227**, 309-317.

57. Brown, S. J., Hilgenfeld, R. B. and Denell, R. E. 1994, The beetle Tribolium castaneum has a fushi tarazu homolog expressed in stripes during segmentation. *Proc Natl Acad Sci U S A*, **91**, 12922-12926.

58. Wang, B., Shao, Z. Q., Xu, Y., et al. 2011, Optimal codon identities in bacteria: implications from the conflicting results of two different methods. *PLoS One*, **6**, e22714.

59. Whittle, C. A., Kulkarni, A. and Extavour, C. G. 2019, Absence of a faster-X effect in beetles (Tribolium, Coleoptera). *BioRxiv; https://doi.org/10.1101/754903*

60. Kearse, M., Moir, R., Wilson, A., et al. 2012, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647-1649.

61. Puigbo, P., Bravo, I. G. and Garcia-Vallve, S. 2008, CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct*, **3**, 38.

62. Whittle, C. A., Malik, M. R. and Krochko, J. E. 2007, Gender-specific selection on codon usage in plant genomes. *BMC Genomics*, **8**, 169-179.

63. Peden, J. F. 1999, Analysis of codon usage. University of Nottingham.

64. Mugal, C. F., von Grunberg, H. H. and Peifer, M. 2009, Transcription-induced mutational strand bias and its effect on substitution rates in human genes. *Mol Biol Evol*, **26**, 131-142.

65. Beletskii, A. and Bhagwat, A. S. 1996, Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli. *Proc Natl Acad Sci U S A*, **93**, 13919-13924.

66. Guo, X., Bao, J. and Fan, L. 2007, Evidence of selectively driven codon usage in rice: implications for GC content evolution of Gramineae genes. *FEBS Lett*, **581**, 1015-1021.

67. Lowe, T. M. and Chan, P. P. 2016, tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res*, **44**, W54-57.

68. Mi, H., Muruganujan, A. and Thomas, P. D. 2013, PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*, **41**, D377-386.

69. Coghlan, A. and Wolfe, K. H. 2000, Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. *Yeast*, **16**, 1131-1145.

70. Polak, P., Querfurth, R. and Arndt, P. F. 2010, The evolution of transcription-associated biases of mutations across vertebrates. *BMC Evol Biol*, **10**, 187.

71. Schober, P., Boer, C. and Schwarte, L. A. 2018, Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, **126**, 1763-1768.

72. Akoglu, H. 2018, User's guide to correlation coefficients. *Turk J Emerg Med*, **18**, 91-93.

73. Semon, M., Lobry, J. R. and Duret, L. 2006, No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Mol Biol Evol*, **23**, 523-529.

74. Plotkin, J. B., Robins, H. and Levine, A. J. 2004, Tissue-specific codon usage and the expression of human genes. *Proc Natl Acad Sci U S A*, **101**, 12588-12591.

669 75. Pang, Y. L., Abo, R., Levine, S. S. and Dedon, P. C. 2014, Diverse cell stresses induce unique
670 patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy
671 number. *Nucleic Acids Res*, **42**, e170.

672 76. Smith, A. M., Abu-Shumays, R., Akeson, M. and Bernick, D. L. 2015, Capture, Unfolding, and
673 Detection of Individual tRNA Molecules Using a Nanopore Device. *Front Bioeng Biotechnol*, **3**,
674 91.

675 77. Sorensen, M. A. 2001, Charging levels of four tRNA species in Escherichia coli Rel(+) and Rel(-
676) strains during amino acid starvation: a simple model for the effect of ppGpp on translational
677 accuracy. *J Mol Biol*, **307**, 785-798.

678 78. Zalucki, Y. M. and Jennings, M. P. 2007, Experimental confirmation of a key role for non-
679 optimal codons in protein export. *Biochem Biophys Res Commun*, **355**, 143-148.

680 79. Percudani, R. 2001, Restricted wobble rules for eukaryotic genomes. *Trends Genet*, **17**, 133-135.

681

Table 1. The organism-wide ΔRSCU between high versus low expressed genes (using averaged expression across all four tissue types, the ovaries, testes, GT-females, and GT-males). In addition, the ΔRSCU are shown when high and low expressed genes were determined for each of the four individual tissue types. The primary optimal (Opt.) codons are in bold and have the largest positive and statistically significant ΔRSCU (t-test P<0.05) per amino acid. For the combined four tissue assessment (organism-wide), the primary optimal (Opt.) and non-optimal codons (Non opt.) are shown with X. Cases where relatively plentiful tRNAs match the optimal codon per amino acid are underlined and bold. The wobble anticodons for codons with zero matching tRNA copies are shown (standard anticodon/wobble anticodon according to classical wobble rules; see also[79]).

		Organism-wide RSCU & ΔRSCU (from average expression across all tissues)								**ΔRSCU per Tissue Type (from expression within each tissue)**							
Amino Acid	**Codon**	**High RSCU**	**Low RSCU**	**ΔRSCU**	**P**	**Opt.**	**Non opt.**	**tRNA No.**	**Standard/Wobble**	**ΔRSCU ovaries**	**P**	**ΔRSCU testes**	**P**	**ΔRSCU female**	**P**	**ΔRSCU male**	**P**
Ala	GCT	1.144	1.001	**+0.143**	**	**X**		**14**		+0.109		**+0.136**	*	**+0.179**	**	+0.146	**
Ala	GCC	1.238	1.203	+0.034				0	GGC/AGC	+0.023		+0.020		+0.115	*	**+0.198**	**
Ala	GCA	0.833	0.867	-0.033				2		-0.065	*[a]	-0.008		-0.110	*	-0.175	**
Ala	GCG	0.731	0.899	-0.168	**		**X**	3		-0.047		-0.128	**	-0.163	**	-0.161	**
Arg	CGT	0.919	0.830	+0.089	*			5		+0.082		+0.087	*	-0.043		+0.116	*
Arg	CGC	0.907	1.117	-0.209	**			0	GCG/ACG	-0.204	*	-0.112	**	-0.231	**	-0.091	*
Arg	CGA	0.946	1.212	-0.265	**		**X**[b]	4		-0.304	**	-0.143	**	-0.189	**	-0.282	**
Arg	CGG	0.650	0.941	-0.290	**		**X**	0	CCG/UCG	-0.235	**	-0.263	**	-0.195	**	-0.272	**
Arg	AGA	1.401	0.990	**+0.411**	**	**X**		**3**		**+0.393**	**	**+0.341**	**	**+0.415**	**	**+0.276**	**
Arg	AGG	1.096	0.801	+0.295	**			3		+0.367	**	+0.247	**	+0.250	**	+0.287	**
Asn	AAT	1.030	0.997	+0.033				0	AUU/GUU	+0.039		+0.041	*[a]	+0.012		-0.027	
Asn	AAC	0.955	0.864	**+0.091**	**	**X**		**5**		**+0.071**	*	**+0.067**	*	**+0.066**	*	**+0.115**	**
Asp	GAT	1.002	0.938	**+0.063**	*	**X**		0	AUC/GAC[e]	**+0.115**	**	**+0.084**	**	**+0.070**	*	+0.031	
Asp	GAC	0.942	0.964	-0.021			**X**[c]	10		-0.035	*[a]	-0.008		-0.026		+0.000	
Cys	TGT	0.986	0.854	**+0.131**	**	**X**		0	ACA/GCA[e]	**+0.204**	**	**+0.130**	**	**+0.155**	**	**+0.103**	**
Cys	TGC	0.802	0.827	-0.025				3		-0.026		+0.028		-0.029	*	-0.006	
Gln	CAA	1.179	1.098	**+0.081**	*	**X**		**5**		+0.043		**+0.089**	*	**+0.099**	*	**+0.064**	*[a]

Gln	CAG	0.758	0.785	-0.026				3		+0.026		+0.002		-0.053		-0.015		
Glu	GAA	1.236	1.110	**+0.125**	**	**X**		**8**		**+0.093**	*	**+0.101**	**	**+0.100**	*	**+0.086**	*	
Glu	GAG	0.733	0.767	-0.034				5		-0.006		+0.002		-0.036		-0.024		
Gly	GGT	0.918	0.801	**+0.116**	**	**X**		0	ACC/GCC[e]	+0.138	*	**+0.133**	**	+0.046		**+0.063**	*	
Gly	GGC	1.017	1.122	-0.104	*		**X**	8		-0.109		-0.077	*	-0.070		-0.001		
Gly	GGA	1.124	1.201	-0.077	*[a]		**X**[b]	15		-0.137	**	-0.071	**	-0.017	*	-0.070	**	
Gly	GGG	0.859	0.792	+0.066				1		**+0.171**	**	+0.072	*	**+0.059**	*	+0.043		
His	CAT	0.840	0.817	+0.023				0	AUG/GUG	+0.055		+0.032		+0.016		-0.017		
His	CAC	1.014	0.978	+0.036				7		**+0.067**	*[a]	+0.054		**+0.053**	*	**+0.084**	*	
Ile	ATT	1.359	1.278	**+0.081**	*	**X**[d]		**7**		**+0.121**	**	**+0.051**	*	**+0.115**	*	+0.033	*[a]	
Ile	ATC	1.024	0.941	**+0.083**	*	**X**[d]		0	GAU/AAU	+0.005		+0.078		+0.012		**+0.165**	**	
Ile	ATA	0.578	0.661	-0.083	*		**X**	2		-0.048		-0.057		-0.071		-0.141	**	
Leu	TTA	0.999	1.127	-0.128	*		**X**	2		-0.087	*	-0.095	*	-0.121	*	-0.211	**	
Leu	TTG	1.794	1.336	**+0.458**	**	**X**		**4**		**+0.409**	**	**+0.391**	**	**+0.339**	**	**+0.444**	**	
Leu	CTT	0.901	0.998	-0.096	*			5		-0.139	**	-0.082	*	-0.003		-0.053	*[a]	
Leu	CTC	0.877	0.926	-0.049				0	GAG/AAG	-0.081		-0.033		-0.042		+0.036		
Leu	CTA	0.492	0.561	-0.068	*			2		+0.023		-0.075	*[a]	-0.003		-0.073	*	
Leu	CTG	0.900	1.008	-0.107	*			2		-0.093		-0.092	*	-0.158	**	-0.108	*	
Lys	AAA	1.272	1.273	-0.000				6		-0.019		-0.005		**+0.068**	*	-0.011		
Lys	AAG	0.728	0.654	**+0.074**	*	**X**		**5**		**+0.075**	*	**+0.067**	*	-0.020		**+0.058**	*	
Phe	TTT	1.058	1.042	+0.015				1		+0.073		+0.003		+0.016		-0.092	**	
Phe	TTC	0.916	0.850	**+0.065**	*	**X**		**5**		+0.015		**+0.076**	*	+0.034		**+0.160**	**	
Pro	CCT	0.904	0.785	+0.119	*			7		**+0.126**	*	+0.092	*	+0.097	*	+0.076	*	
Pro	CCC	1.090	0.917	**+0.172**	**	**X**		0	GGG/AGG[e]	+0.064	*	**+0.131**	*	**+0.163**	*	**+0.264**	**	
Pro	CCA	1.044	1.014	+0.029			**X**[c]	13		+0.166		+0.021		+0.063		-0.021		
Pro	CCG	0.889	1.102	-0.213	**		**X**	1		-0.220	**	-0.140	**	-0.249	**	-0.237	**	
Ser	TCT	0.849	0.732	+0.116	*			4		+0.084		+0.073		+0.061		+0.026		
Ser	TCC	0.894	1.056	-0.162	**			0	GGA/AGA	-0.137		-0.152	**	-0.051		-0.010		
Ser	TCA	1.059	0.977	+0.082	*[a]			2		+0.114		+0.072	*	-0.017		+0.001		
Ser	TCG	1.128	1.231	-0.103	*			2		-0.066		-0.023		-0.101	*	-0.012		

Ser	AGT	1.149	0.922	**+0.226**	**	**X**		0	ACU/GCU[e]	**+0.218**	**	**+0.197**	**	**+0.268**	**	**+0.137**	**
Ser	AGC	0.900	1.039	-0.138	*		**X**	3		-0.156	*	-0.156	**	-0.125	**	-0.105	**
Thr	ACT	1.107	0.884	**+0.222**	**	**X**		**5**		**+0.199**	**	**+0.266**	**	**+0.188**	**	**+0.207**	**
Thr	ACC	1.032	1.003	+0.029				0	GGU/AGU	+0.026		-0.026	*[a]	+0.148	*	+0.178	**
Thr	ACA	1.001	1.013	-0.012				3		-0.027		-0.059		-0.076		-0.136	**
Thr	ACG	0.812	1.006	-0.194	**		**X**	2		-0.129	*	-0.113	*	-0.213	**	-0.211	**
Tyr	TAT	0.819	0.881	-0.062	*			0	AUA/GUA	+0.040		-0.018	*	-0.041	*	-0.080	**
Tyr	TAC	1.096	0.898	**+0.197**	**	**X**		**13**		**+0.123**	*	**+0.162**	**	**+0.156**	**	**+0.210**	**
Val	GTT	1.262	1.133	**+0.129**	*	**X**		**7**		**+0.124**	*	**+0.119**	*	**+0.157**	*	**+0.101**	*
Val	GTC	0.976	0.999	-0.022				0	GAC/AAC	-0.017		-0.034	*[a]	-0.027		+0.053	
Val	GTA	0.552	0.625	-0.073	*		**X**	5		-0.047	*	-0.064	*	-0.019		-0.082	*
Val	GTG	1.156	1.140	+0.015				3		+0.020		+0.032		-0.055		-0.041	

**P<0.001, *P<0.05 and ≥0.001; [a] P-values are between 0.05 and 0.1 and thus is considered a putative optimal codon; [b] Secondary non-optimal codon with relatively high matching tRNA count; [c] While not having a statistically significant negative ΔRSCU, the codon is not optimal and is notable by its high tRNA count; [d] Both codons are optimal codons at nearly the same level; [e] Codon has Opt-codon$_{wobble}$ status.

Table 2. Examples of functions of the subset of highly transcribed genes that have elevated usage of non-optimal codons (RSCU>1.5) with substantial accepting tRNA gene counts (≥4). The genes are candidates to be translationally regulated by the degree of non-optimal codon usage.

High GGC Usage for Gly (RSCU>1.5)	
Gene Functions	
Ribosomal protein genes	
TC006109	14-3-3 protein epsilon-like Protein
TC011123	40S ribosomal protein S13-like Protein
TC008667	40S ribosomal protein S20-like Protein
TC005084	40S ribosomal protein S26
TC010830	40S ribosomal protein S6
TC009214	40S ribosomal protein S7
TC014757	40S ribosomal protein S8
TC016306	60S acidic ribosomal protein P0
TC010413	60S acidic ribosomal protein P1-like Protein
TC015013	60S acidic ribosomal protein P2-like Protein
TC013536	60S ribosomal protein L17-like Protein
TC007932	60S ribosomal protein L21-like Protein
TC013168	60S ribosomal protein L4-like Protein
TC030666	60S ribosomal protein L6-like Protein
TC011182	60S ribosomal protein L7a-like Protein
Olfactory	
TC007741	Odorant binding protein 12
TC010070	Odorant binding protein C06
TC008681	Chemosensory protein 1
Stress-response	
TC004948	Peroxiredoxin 1-like Protein
TC014929	Peroxiredoxin 1-like Protein
Uncharacterized Proteins (N=50)	
High GTA usage for Val	
Cytoskeletal	
TC001574	Cofilin/actin-depolymerizing factor homolog-like Protein
TC033072	profilin
p53 related	
TC034594	Cell death-inducing p53-target protein 1-like protein
Ribosomal protein genes (N=0)	
Uncharacterized Proteins (N=15)	
High CGA usage for Arg	
Olfactory	
TC010070	Odorant binding protein C06; TcOBP7M ortholog
TC030421	Odorant receptor 305; Or305; ortholog
TC008681	Chemosensory protein 1; TcCSP7K; ortholog
p53 related	
TC034594	Cell death-inducing p53-target protein 1-like protein
Cytoskeletal	
TC007700	Tubulin-specific chaperone cofactor E-like protein
TC009721	Microtubule-protein RP/EB family member 1
TC012270	Troponin C, isoform 1-like Protein
TC033072	Profilin
TC001942	Putative dynactin subunit 2-like Protein (Fragment)
Ribosomal protein genes (N=0)	
Uncharacterized Proteins (N=15)	

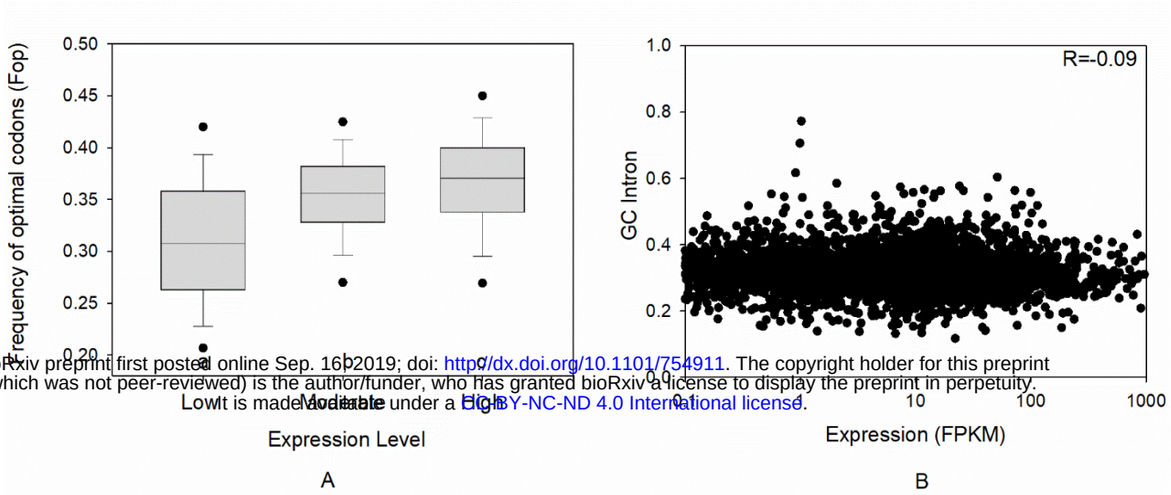

Figure 1. A. The frequency of optimal codons (Fop) across all 16,434 genes studied in *T. castaneum*. Genes are categorized into low (lowest 5%, FPKM<0.013), moderate (5 to 95%) and high (top 5%) transcription (FPKM>103) groups based on average expression across all four tissue types (testes, ovaries, GT-males, GT-females). Different letters below bars indicate a statistically significant difference using Ranked ANOVA and Dunn's paired contrasts (P<0.05). **B.** The GC content of introns with respect to the expression level per gene (Spearman's Ranked R is shown). Values are shown for all genes with introns >50bp (N=5,143).

SUPPLEMENTARY MATERIAL

Evidence of multifaceted functions of codon usage in translation within the model beetle *Tribolium castaneum*

Carrie A. Whittle, Arpita Kulkarni, Cassandra G. Extavour

Table S1. The number of RNA-seq reads for each tissue-type in the present study [1]. RNA-seq data are shown before and after adapter and quality trimming with BBDuk (https://jgi.doe.gov/data-and-tools/bbtools/). The Short Read Archive (SRA) Biosample identifiers are also shown (https://www.ncbi.nlm.nih.gov/sra).

Tissue Sample[a]	**No. of Reads**		**SRA Biosample ID**
	Before trimming	**After trimming**	
Tribolium castaneum			
Testes sample 1	18,006,255	17,995,655	SAMN12702873
Ovary sample 1	39,140,493	39,122,050	SAMN12702874
GT-male sample 1	25,630,261	25,609,723	SAMN12702875
GT-female sample 1	41,513,717	41,472,348	SAMN12702876
Testes sample 2	24,795,583	24,787,238	SAMN12702877
Ovary sample 2	22,306,622	22,286,961	SAMN12702878
GT-male sample 2	62,781,001	62,712,242	SAMN12702879
GT-female 2	52,275,340	52,211,149	SAMN12702880

[a] Reads were obtained from for two RNA-seq runs of each biological sample.

Table S2. Comparison of the primary optimal codon list generated using the organism-wide analyses of high and low expressed genes in the present study (ΔRSCU) to optimal codons obtained using the correlation method in Williford and Demuth (2012), which defined up to three optimal codons per amino acid. Cases wherein the present primary optimal codon matched an optimal codon (at any level) identified under the correlation approach are indicated in the right-most column. Optimal codons defined in each study are in bold and underlined.

ΔRSCU Method Herein			Correlation Method [2]					Same optimal codon
Amino Acid	Codon	RSCU (All Tissues)	Amino Acid	Codon	Female RT	Male RT	Female & Male whole body	
Ala	**GCT**	**+0.143**	Ala	GCT	−0.028	−0.014	0.005	NO*
Ala	GCC	+0.034	Ala	**GCC**	0.176	0.16	0.184	
Ala	GCA	-0.033	Ala	GCA	−0.17	−0.15	−0.155	
Ala	GCG	-0.168	Ala	**GCG**	0.078	0.061	0.017	
Arg	CGT	+0.089	Arg	CGT	0.016	0.008	0.0002	
Arg	CGC	-0.209	Arg	**CGC**	0.068	0.075	0.052	
Arg	CGA	-0.265	Arg	CGA	−0.15	−0.154	−0.13	
Arg	CGG	-0.290	Arg	CGG	−0.012	−0.029	−0.038	
Arg	**AGA**	**+0.411**	Arg	AGA	−0.006	0.012	0.031	NO
Arg	AGG	+0.295	Arg	**AGG**	0.23	0.225	0.209	
Asn	AAT	+0.033	Asn	AAT	−0.112	−0.1	−0.135	
Asn	**AAC**	**+0.091**	Asn	**AAC**				**YES**
Asp	**GAT**	**+0.063**	Asp	GAT	−0.028	−0.023	−0.043	NO
Asp	GAC	-0.021	Asp	GAC				
Cys	**TGT**	**+0.131**	Cys	TGT	−0.014	0.009	−0.011	NO
Cys	TGC	-0.025	Cys	TGC				
Gln	**CAA**	**+0.081**	Gln	CAA	−0.101	−0.075	−0.064	NO
Gln	CAG	-0.026	Gln	**CAG**				
Glu	**GAA**	**+0.125**	Glu	GAA	−0.156	−0.137	−0.108	NO
Glu	GAG	-0.034	Glu	**GAG**				
Gly	**GGT**	**+0.116**	Gly	GGT	0.008	0.022	0.03	NO
Gly	GGC	-0.104	Gly	**GGC**	0.095	0.069	0.067	
Gly	GGA	-0.077	Gly	GGA	−0.196	−0.203	−0.145	
Gly	GGG	+0.066	Gly	**GGG**	0.196	0.205	0.157	
His	CAT	+0.023	His	CAT	−0.043	−0.054	−0.052	
His	**CAC****	+0.036	His	**CAC**				**YES****
Ile	ATT	+0.081	Ile	ATT	−0.089	−0.059	−0.087	
Ile	**ATC**	**+0.083**	Ile	ATC	0.148	0.137	0.171	**YES**
Ile	ATA	-0.083	Ile	ATA	−0.043	−0.065	−0.078	
Leu	TTA	-0.128	Leu	TTA	−0.07	−0.096	−0.121	
Leu	**TTG**	**+0.458**	Leu	**TTG**	0.182	0.237	0.188	**YES**
Leu	CTT	-0.096	Leu	CTT	−0.123	−0.111	−0.081	
Leu	CTC	-0.049	Leu	**CTC**	0.076	0.05	0.084	

Pro	CCG	-0.213	Pro	CCG	0.015	0.001	−0.004	
Ser	TCT	+0.116	Ser	TCT	−0.06	−0.071	−0.053	
Ser	TCC	-0.162	Ser	TCC	0.015	−0.001	0.037	
Ser	TCA	+0.082	Ser	TCA	−0.041	−0.039	−0.048	
Ser	TCG	-0.103	Ser	**TCG**	0.103	0.09	0.073	
Ser	**AGT**	**+0.226**	Ser	**AGT**	0.071	0.108	0.075	**YES**
Ser	AGC	-0.138	Ser	**AGC**	0.056	0.045	0.06	
Thr	**ACT**	**+0.222**	Thr	ACT	0.004	0.031	0.025	NO
Thr	ACC	+0.029	Thr	**ACC**	0.111	0.097	0.142	
Thr	ACA	-0.012	Thr	ACA	−0.12	−0.119	−0.123	
Thr	ACG	-0.194	Thr	**ACG**	0.088	0.079	0.039	
Tyr	TAT	-0.062	Tyr	TAT	−0.094	−0.107	−0.124	
Tyr	**TAC**	**+0.197**	Tyr	**TAC**				**YES**
Val	**GTT**	**+0.129**	Val	GTT	−0.049	−0.031	−0.038	NO
Val	GTC	-0.022	Val	**GTC**	0.09	0.077	0.106	
Val	GTA	-0.073	Val	GTA	−0.057	−0.069	−0.058	
Val	GTG	+0.015	Val	**GTG**	0.104	0.112	0.081	

* The same optimal codon GCC was found for ovaries and testes when examined individually in the present study.

** The CAC codon is statistically significantly optimal for three of four tissues herein, but not in the summary analyses of all pooled tissues. It is included in comparison of the present optimal codons to the correlation method.

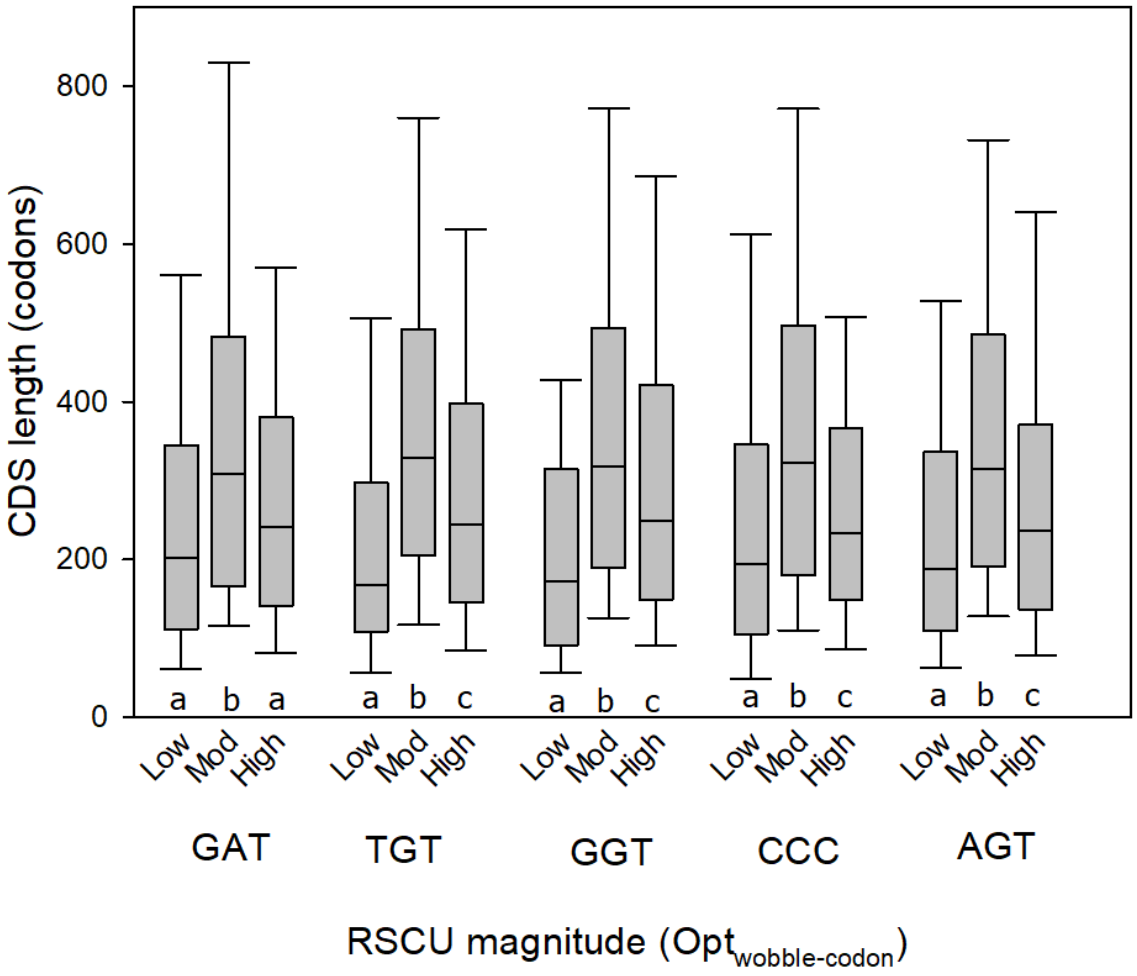

Figure S1. The relative use of codons with Opt-codon$_{wobble}$ status (GAT, TGT, GGT, CCC and AGT) in highly expressed genes with respect to CDS length. Different letters below each set of three bars (per codon) indicate a statistically significant difference using Ranked ANOVA and Dunn's paired contrasts ($P<0.05$). The 822 highly expressed genes were divided into three equal sized classes of RSCU values (low, moderate (mod), high) for each codon.

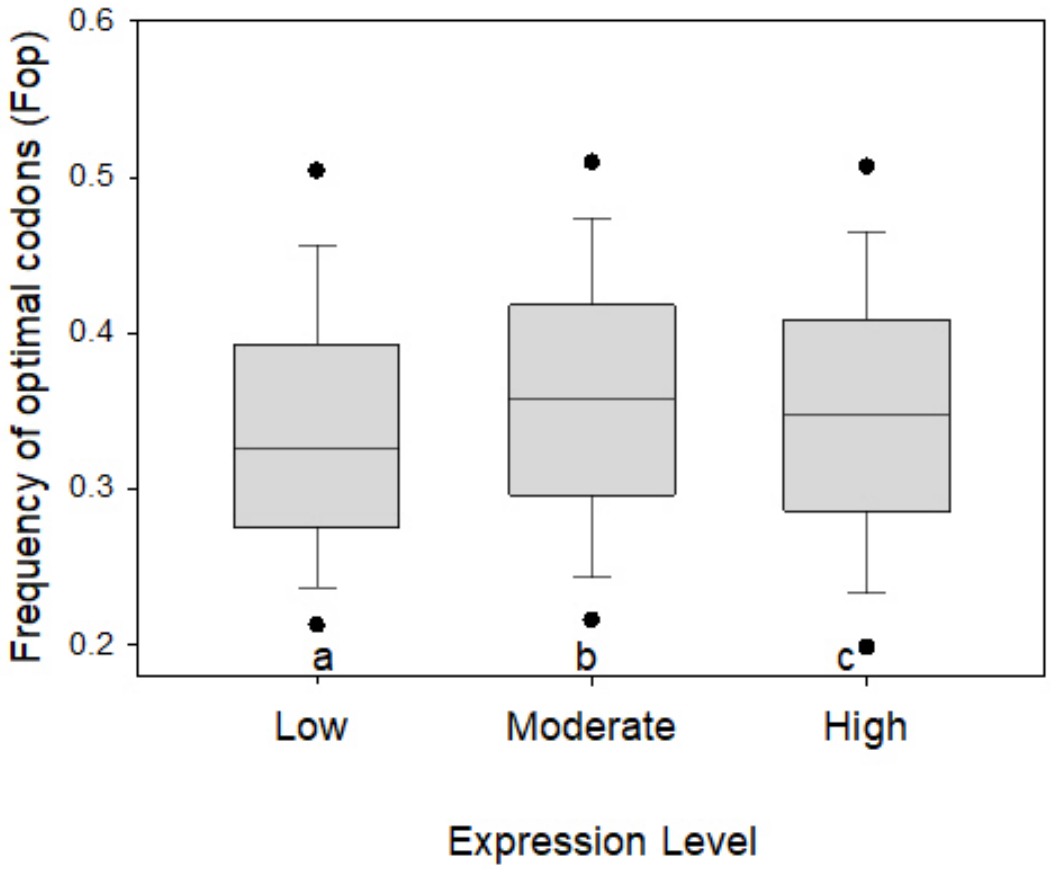

Figure S2. The frequency of optimal codons (Fop) across all genes studied in *T. castaneum* when using the primary optimal codons identified in Williford and Demuth.[2] Genes are categorized into low (lowest 5%, FPKM<0.013), moderate (5 to 95%) and high (top 5%) transcription groups (FPKM>103) based on average expression across all four tissue types (testes, ovaries, GT-males, GT-females). Different letters below bars indicate a statistically significant difference using Ranked ANOVA and Dunn's paired contrasts (<0.05).

Supplementary Text File S1: Protein length and $Opt_{\uparrow tRNA}$ status

For the beetles studied herein, we found that the use of Opt-$codon_{wobble}$ codons was connected to protein length. Specifically, for the 822 highly transcribed genes in this organism (top 5%), we ranked the RSCU for each of the five codons with Opt-$codon_{wobble}$ status, and genes were then binned into three equal sized categories (N=274 genes each) based on the relative magnitude of RSCU (low, moderate, and high). By definition as an optimal codon, each of these five codons had elevated RSCU in the highly expressed genes (as compared to low expressed genes, Table 1). However, within the highly transcribed gene set, we found that the bin containing moderate RSCU values were consistent linked to longer CDS than those with the lowest or highest RSCU values for each of the five Opt-$codon_{wobble}$ codons, namely GAT, TGT, GGT, CCC and AGT (Ranked ANOVA and Dunn's paired contrast $P<0.05$, Fig. S1). Thus, the highly transcribed CDS encoding long proteins, appear to be connected to a specific frequency of Opt-$codon_{wobble}$ codons, which may play a role in their translation. This may possibly comprise a mechanism to ensure a balance between high translation rates (ensured by moderate rather than highest usage of Opt-$codon_{wobble}$ codons) and allowing intermittent pausing during translation for accurate protein synthesis and/or protein folding (ensured by their moderate, rather than low, usage) of CDS encoding long proteins.

Supplementary References

1. Whittle, C. A., Kulkarni, A. and Extavour, C. G. 2019, Absence of a faster-X effect in beetles (Tribolium, Coleoptera). *BioRxiv; https://doi.org/10.1101/754903*

2. Williford, A. and Demuth, J. P. 2012, Gene expression levels are correlated with synonymous codon usage, amino acid composition, and gene architecture in the red flour beetle, Tribolium castaneum. *Mol Biol Evol*, **29**, 3755-3766.