

1 **Adaptation of codon and amino acid use for translational functions**
2 **in highly expressed cricket genes**

3

4 Authors: Carrie A. Whittle¹, Arpita Kulkarni¹, Nina Chung¹, Cassandra G. Extavour^{1,2}

5

6 1. Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity
7 Avenue, Cambridge MA 02138, USA

8 2. Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue,
9 Cambridge MA

10

11

12 Corresponding author: Cassandra G. Extavour

13 Email: extavour@oeb.harvard.edu

14 Phone: (617) 496 1935

15 Fax: (617) 496 9507

16

17 **Abstract**

18

19 **Background**

20 For multicellular organisms, much remains unknown about the dynamics of synonymous codon
21 and amino acid use in highly expressed genes, including whether their use varies with expression
22 in different tissue types and sexes. Moreover, specific codons and amino acids may have
23 translational functions in highly transcribed genes, that largely depend on their relationships to
24 tRNA gene copies in the genome. However, these relationships and putative functions are poorly
25 understood, particularly in multicellular systems.

26

27 **Results**

28 Here, we rigorously studied codon and amino acid use in highly expressed genes from
29 reproductive and nervous system tissues (male and female gonad, somatic reproductive system,
30 brain, ventral nerve cord, and male accessory glands) in the cricket *Gryllus bimaculatus*. We
31 report an optimal codon, defined as the codon preferentially used in highly expressed genes, for
32 each of the 18 amino acids with synonymous codons in this organism. The optimal codons were
33 largely shaped by selection, and their identities were mostly shared among tissue types and both
34 sexes. However, the frequency of optimal codons was highest in gonadal genes. Concordant with
35 translational selection, a majority of the optimal codons had abundant matching tRNA gene
36 copies in the genome, but sometimes obligately required wobble tRNAs. We suggest the latter
37 may comprise a mechanism for slowing translation of abundant transcripts, particularly for cell-
38 cycle genes. Non-optimal codons, defined as those least commonly used in highly transcribed
39 genes, intriguingly often had abundant tRNAs, and had elevated use in a subset of genes with
40 specialized functions (gametic and apoptosis genes), suggesting their use promotes the
41 upregulation of particular mRNAs. In terms of amino acids, we found evidence suggesting that
42 amino acid frequency, tRNA gene copy number, and amino acid biosynthetic costs
43 (size/complexity) had all interdependently evolved in this insect model, potentially for
44 translational optimization.

45

46 **Conclusions**

47 Collectively, the results strongly suggest that codon use in highly expressed genes, including
48 optimal, wobble, and non-optimal codons, and their tRNAs abundances, as well as amino acid
49 use, have been adapted for various functional roles in translation within this cricket. The effects
50 of expression in different tissue types and the two sexes are discussed.

51

52 **Keywords:** Codon, amino acid, tissue-type, translational selection, regulation, tRNAs

53

54 **Background**

55 Synonymous codons in protein-coding genes are not used randomly [1]. The preferential
56 use of synonymous codons per amino acid in highly transcribed genes, often called optimal
57 codons, has been observed in diverse organisms including bacteria, fungi, plants and animals [2-
58 14], including insects such as flies, mosquitoes, beetles and crickets [10, 11, 13, 15, 16]. When
59 optimal codons co-occur with a high count of iso-accepting tRNA gene copies in the genome,
60 which reflects an organism's tRNA abundance [3-5, 12, 17-20], it suggests a history of selection
61 favoring translational optimization [1, 3, 5, 12, 13, 20-25]. In multicellular organisms, unlike
62 unicellular systems, genes can be expressed at different levels among tissue types and between
63 the two sexes [16, 26-29]. Thus, in these organisms, codon use may be more complex, given that
64 it is plausible that optimal codons may depend on the tissue type or sex in which a gene is
65 expressed [11, 16, 22, 30, 31], and codon use could feasibly adapt to local tissue-dependent
66 tRNA populations [30, 32, 33]. However, only minimal data are currently available about
67 whether and how codon use varies with high expression in different tissue types and between the
68 two sexes in multicellular organisms.

69 The limited data that are available suggest that codon use varies among genes transcribed
70 in different tissues. We recently found, for example, that some optimal codons of highly
71 transcribed genes differed among males and females for the testis, ovaries, gonadectomized-
72 males and gonadectomized females, which may suggest adaptation of codon use to local tRNA
73 populations in the beetle *T. castaneum* [16]. In addition, a study in *Drosophila melanogaster*
74 showed that certain codons were preferentially used in the testis (CAG (Gln), AAG (Lys), CCC
75 (Pro), and CGU (Arg)) as compared to other tissues such as the midgut, ovaries, and salivary
76 glands, a result that was taken as support for the existence of tissue-specific tRNA populations
77 [32] (see also an analysis of codon bias by [31]). Similar patterns of tissue-related use of specific
78 codons have been inferred in humans [33, 34] and the plants *Arabidopsis thaliana* and *Oryza*
79 *sativa* [30, 35]. Given the limited scope of organisms studied to date, however, further research
80 is needed to determine whether the codon use varies among tissues across a broader scale of
81 organisms. Tissues that are of particular importance for research include the gonads, which are
82 key to reproductive success, and the brain, wherein the transcribed genes are apt to regulate male
83 and female sexual behaviors [36-38]. Translational optimization of highly transcribed genes in
84 these tissues may be particularly significant for an organism's fitness.

85 While much of the focus on codon use in an organism’s highly expressed genes to date
86 has centered on optimal codons [3, 5, 7, 12, 13, 16, 21-25, 39-41], and whether they have
87 abundant matching tRNAs that may improve translation [3, 12, 13, 20-24], growing evidence
88 suggests that other, less well studied, types of codon statuses could also play important
89 translational roles [42-44]. In particular, even for codons that are not optimal, the supply-demand
90 relationship between codons and tRNA abundances may regulate translation rates, possibly
91 affecting protein functionality and abundance [42, 45-47]. For example, *in vivo* experimental
92 research has shown that genes using codons requiring wobble tRNAs, which imprecisely match a
93 codon at the third nucleotide site, exhibit slowed movement of ribosomes along mRNAs [42, 48,
94 49]. Similarly, non-optimal codons, defined as those codons that are least commonly used in
95 highly transcribed genes (or sometimes defined as “rare” codons), particularly those non-optimal
96 codons with few or no tRNAs in the cellular tRNA pool [16], may decelerate translation and
97 thereby prevent ribosomal jamming [19] and also allow proper co-translational protein folding
98 [44, 50-53]. In this regard, wobble codons, and non-optimal codons with few matching tRNA
99 gene copies in the genome, may have significant translational roles, including roles in slowing
100 translation.

101 In contrast to non-optimal codons that have few tRNAs, some evidence has emerged
102 suggesting non-optimal codons may sometimes have abundant tRNAs, a relationship that may
103 act to improve translation of specific gene mRNAs [16, 45]. For instance, in yeast
104 (*Saccharomyces cerevisiae*), rare genomic codons exhibit enhanced use in stress genes, and
105 tRNAs matching these codons have been found to be upregulated in response to stressful
106 conditions, allowing improvement of their translation levels without any change in transcription
107 rates [45]. In the red flour beetle, we recently reported that some non-optimal codons have
108 abundant matching tRNA genes in the genome [16], and these codons are concentrated in a
109 subset of highly transcribed genes with specific, non-random biological functions (e.g., olfactory
110 or stress roles), which may together allow preferential translation of mRNAs of those particular
111 genes [16]. Accordingly, given these findings, further studies of codon use patterns in highly
112 expressed genes of multicellular organisms should expand beyond the focus on optimal codons
113 *per se* [2, 3, 7-9, 12, 21, 39, 41], and explore the use and possible translational functions of non-
114 optimal codons, distinguishing between those that have few and plentiful tRNAs, as well as the
115 use of wobble codons [16].

116 While the investigation of amino acid use remains uncommon in multicellular organisms,
117 the available sporadic studies suggest an association between amino acid use and gene
118 expression level [21, 54, 55]. In insects, for example, an assessment of the biosynthetic costs of
119 amino acid synthesis (size/complexity score for each of 20 amino acids as quantified by Dufton
120 [56]) has shown that those amino acids with low costs tend to be more commonly used in genes
121 with high transcription levels in the beetle *T. castaneum* [21]. Further, genome-wide studies in
122 other arthropod models such as spiders (*Parasteatoda tepidariorum*) [55], and the study of
123 partial available transcriptomes from milkweed bugs (*Oncopeltus fasciatus*), an amphipod
124 crustacean (*Parhyale hawaiiensis*) and crickets (*Gryllus bimaculatus*, using a single
125 ovary/embryo dataset in this system) [10], were suggestive of the hypothesis that evolution may
126 have typically favored a balance between minimizing the amino acid costs for production of
127 abundant proteins with the need for certain (moderate cost) amino acids to ensure proper protein
128 function (protein stability and/or functionality) [54]. Moreover, it has been found that amino acid
129 use is correlated to their tRNA gene copy numbers in beetles [21], and in some other eukaryotes
130 [17], a relationship that may be stronger in highly transcribed genes [17]. Thus, these various
131 patterns raise the possibility of adaptation of amino acid use for translational optimization in
132 multicellular organisms [17, 21, 55]. At present, further data is needed on amino acid use in
133 highly expressed genes in multicellular systems, that include consideration of tRNA gene
134 number, biosynthetic costs, and expression in different tissue types.

135 An emerging model system that provides opportunities for further deciphering the
136 relationships between gene expression and codon and amino acid use is the two-spotted cricket
137 *Gryllus bimaculatus*. Within insects, *Gryllus* is a hemimetabolous genus (Order Orthoptera) and
138 has highly diverged from the widely studied model insect genus *Drosophila* (Order Diptera) [57,
139 58]. *G. bimaculatus* comprises a model for investigations in genetics [59, 60], germ line
140 formation and development [61-63] and for molecular evolutionary biology [10, 64]. In the
141 present study, we rigorously assess codon and amino acid use in highly transcribed genes of *G.*
142 *bimaculatus* using its recently available annotated genome [65] and large-scale RNA-seq data
143 from tissues of the male and female reproductive and nervous systems [64]. From our analyses,
144 we demonstrate that optimal codons, those preferentially used in highly expressed genes, occur
145 in this organism, are largely shaped by selection pressures, and are nearly identical across
146 tissues. Based on analyses of codon and tRNA gene copy relationships, we find that a majority of

147 optimal codons have abundant tRNAs, which is consistent with translational optimization in this
148 species. However, some optimal codons obligately require the use of wobble tRNAs, which may
149 act to slow translation, including for cell-cycle genes. Moreover, non-optimal codons, those
150 codons rarely used in highly expressed genes, rather than usually having few tRNAs, often have
151 abundant tRNAs, and thus may provide a system to upregulate the translation of specific mRNAs
152 (for example, apoptosis gonadal genes), as has been proposed in yeast and beetles [16, 45].
153 Finally, with respect to amino acids, we find evidence to suggest that amino acid frequency,
154 tRNA gene copy number, and amino acid biosynthetic costs have all interdependently evolved in
155 this taxon, possibly for translational optimization.

156

157 **Results and Discussion**

158 For our study, codon and amino acid use in *G. bimaculatus* was assessed using genes
159 from its recently available annotated genome [65]. We included all 15,539 *G. bimaculatus*
160 protein-coding genes (CDS, longest CDS per gene) that had a start codon and were >150bp.
161 Gene expression was assessed using RNA-seq data from four adult male and female tissue types,
162 the gonad (testis for males, ovaries for females), somatic reproductive system (for males this
163 includes the pooled vasa deferentia, seminal vesicle and ejaculatory duct and for females
164 includes the spermathecae, common oviduct, and bursa), brain and ventral nerve cord
165 (Additional file 1: Table S1; [64]). The male accessory glands were included for study, but were
166 separated from the other male reproductive system to prevent overwhelming, or skewing, the
167 types of transcripts detected in the former tissues [64]. The trimmed reads in Additional file 1:
168 Table S1 were mapped to the 15,539 annotated *G. bimaculatus* genes independently for each of
169 the nine tissue types under study and the expression level, or FPKM, was determined per gene.

170 To identify the optimal and non-optimal codons in *G. bimaculatus*, we compared codon
171 use in highly versus lowly expressed genes [2, 7, 9, 10, 15, 16, 39, 66, 67]. For each CDS, the
172 relative synonymous codon usage (RSCU) was determined for all codons for each amino acid
173 with synonymous codons, whereby RSCU values >1 and <1 respectively indicate greater and
174 lower use of a synonymous codon than that expected under equal codon use, and elevated values
175 of codons for each amino acid indicate more frequent usage [18]. The $\Delta\text{RSCU} = \text{RSCU}_{\text{Mean Highly Expressed CDS}} - \text{RSCU}_{\text{Mean Low Expressed CDS}}$
176 was used to define the primary optimal codon as the codon
177 with the largest positive and statistically significant ΔRSCU value per amino acid [2, 7, 9, 10, 15,

178 16, 39]. The primary non-optimal codon was defined as the codon with the largest negative and
179 statistically significant Δ RSCU value per amino acid [16]. In the following sections, we first
180 thoroughly describe the optimal codons identified in this cricket species, including an assessment
181 of the variation in expression among tissue types, and the role of selection versus mutation in
182 shaping the optimal codons. Subsequently, we thoroughly evaluate the relationships between
183 optimal codons and the non-optimal codons and their matching tRNA gene counts in the genome
184 to ascertain plausible functional roles. We then consider the amino acid use and tRNA
185 relationships in highly expressed genes of this taxon.

186

187 **Optimal Codons are Shared Across the Nine Distinct Tissues in *G. bimaculatus***

188 The organism-wide optimal codons were identified for *G. bimaculatus* using Δ RSCU for
189 genes with the top 5% average expression levels across all nine studied tissues (cutoff was 556.2
190 FPKM) versus the 5% of genes with the lowest average expression levels (among all 15,539
191 genes under study) and are shown in Table 1. Based on Δ RSCU we report a primary optimal
192 codon for all of the 18 amino acids with synonymous codons, each of which ended at the third
193 position in an A (A3) or T (T3) nucleotide (boldface and underlined Δ RSCU values, Table 1). As
194 shown in Table 2, the 777 genes in the top 5% average expression category (organism-wide
195 analysis) were enriched for ribosomal protein genes and had mitochondrial and protein folding
196 functions. We found that 14 of the 17 primary optimal codons (one per amino acid) that were
197 previously identified using a partial transcriptome from one pooled tissue sample
198 (embryos/ovaries [10]), were identical to those observed here, marking a strong concordance
199 between studies and datasets (the differences herein were CAA for Gln, TTA for Leu, and AGA
200 for Arg as optimal codons, and the presence of an optimal codon AAA for Lys, which had no
201 optimal codon using previous embryonic/ovary data [10]). Thus, the present analysis using large-
202 scale RNA-seq from nine divergent tissues (Additional file 1: Table S1) and using a complete
203 annotated genome [65] support a strong preference for AT3 codons in this cricket.

204 Importantly, the expression datasets herein (Additional file 1: Table S1) allowed us to
205 conduct an assessment of whether the identity of optimal codons varied with tissue type or sex.
206 As certain data suggest that codon use may be influenced by the tissue in which it is maximally
207 transcribed [16, 30], we examined those genes that exhibited maximal expression (in the top 5%)
208 within each tissue type, that were not in the top 5% for any of the other eight remaining tissue

209 types [16, 30], which we refer to as Top5_{One-tissue} (N values as follows, female gonad (274), male-
210 gonad (270), female somatic reproductive system (67), male somatic reproductive system (104),
211 female brain (24), male brain (22); female ventral nerve cord (32), male ventral nerve cord (33),
212 and male accessory glands (162)). We found remarkable consistency among tissues, with nearly
213 all identified optimal codons (largest positive Δ RSCU and $P < 0.05$) ending in A3 and T3 in each
214 tissue (Additional file 1: Table S2). For amino acids with two codons, the organism-wide optimal
215 codon was always optimal across all nine tissues (Additional file 1: Table S2; with possible
216 exceptions for the optimal codons AGG for Arg and CAG for Gln in the male brain; however
217 this had $P > 0.1$, and the N values and thus statistical power was lowest for the male brain;
218 Additional file 1: Table S2). Nonetheless, there was some minor variation among the AT3-ending
219 codons for amino acids with three or more synonymous codons. As an example, for the amino
220 acid Thr, ACT was the optimal codon at the organism-wide level (Table 1) and for five tissues
221 types (male somatic reproductive system, male brain, male ventral nerve cord, female ventral
222 nerve cord, and male accessory glands), while the secondary organism-wide optimal codon ACA
223 (secondary status is based on their magnitude of $+\Delta$ RSCU values) was the primary optimal
224 codon in four other tissues (Additional file 1: Table S2). Thus, for some amino acids there is
225 mild variation in primary and secondary status among tissues of the AT3 codons, which may
226 reflect modest differences in the tRNA abundances among tissues [16, 32]. However, the overall
227 patterns suggest there is remarkably high consistency in the identity of AT3 optimal codons
228 across diverse tissues in this taxon (Additional file 1: Table S2).

229 While tissue-related optimal codons in multicellular organisms have only rarely been
230 studied, the data available from fruit flies, thale cress (*Arabidopsis*), and our recent results from
231 red flour beetles [16, 30, 32] have shown that optimal codons can vary among tissues, which
232 suggests the existence of tissue-specific tRNA pools in those taxa [32]. The results here in *G.*
233 *bimaculatus* thus differ from those in other organisms, and suggest its tRNA pools do not vary
234 substantially with tissue or sex. Future studies using direct quantification of tRNA populations in
235 various tissue types, which is a methodology under refinement and wherein the most effective
236 approaches remain debated [45, 68], will help further affirm whether tRNA populations are
237 largely similar among tissues and sex in this organism. Taken together, the results from this
238 Top5_{One-tissue} analysis, wherein the gene set for each tissue is mutually exclusive of the top 5%
239 expressed genes in any other tissue, suggest that high transcription in even a single tissue type or

240 sex is enough to give rise to the optimal codons in this species. We note nonetheless that while
241 the identity of optimal codons, and thus potentially the relative tRNA abundances, are shared
242 among genes expressed in different tissues, the frequency of optimal codons (Fop) [22] varied
243 among tissue types (Top5_{One-tissue}), suggesting the absolute levels of tRNAs may differ among
244 tissues (see below section “*Fop varies with tissue type and sex*”).

245

246 ***Selective pressure is a primary factor shaping optimal codons***

247 Given that the optimal codons were highly consistent across tissues, to further investigate
248 the potential role of selection in shaping the optimal codons we focused on the organism-wide
249 optimal codons (Table 1). While the elevated use of the specific types of codons in highly
250 expressed genes in Table 1 in itself provides evidence of a history of selection favoring the use
251 of optimized codons in *G. bimaculatus* [2, 7, 9, 10, 15, 16, 66, 67, 69], the putative role of
252 selection can be further evaluated by studying the AT (or GC) content of introns (AT-I), which
253 are thought to largely reflect background neutral mutational pressures on genes, and thus on AT3
254 [16, 66, 70-74]. The *G. bimaculatus* genome contains repetitive A and T rich non-coding DNA
255 [65], including in the introns. Nonetheless, to decipher whether any additional insights might be
256 gained from the introns in *G. bimaculatus* we extracted the introns from the genome and found
257 that 90.5% (N=14,071) of the 15,539 annotated genes had introns suitable for study (≥ 50 bp after
258 trimming). The AT-I content across all genes in this taxon had a median of 0.637, indicating a
259 substantial background compositional nucleotide bias, and differing from the whole gene CDS
260 (median AT for CDS across all sites=0.525). Introns (longest per gene) were nearly two- fold
261 shorter for the most highly (top 5% organism-wide) than lowly (lowest 5%) expressed genes
262 (1.91 fold longer in low than high expressed genes, medians were 5,183 and 2,694bp
263 respectively, MWU-test $P < 0.05$). We speculate that the shorter introns under high expression
264 may comprise a mechanism to minimize transcriptional costs of abundantly produced transcripts
265 in this cricket, as has been suggested in some other species including humans and nematodes
266 [75], and may indicate a history of some non-neutral evolutionary pressures on the length of
267 introns.

268 To further distinguish the role of mutation from selection in shaping AT3 in this cricket,
269 we evaluated the relationship between gene expression (FPKM) and AT-I and AT3. We found
270 that AT-I was positively correlated to gene expression level, with Spearman’s $R = 0.354$, $P < 2 \times 10^{-10}$.

271 7 (across all 14,071 annotated genes with introns). Thus, assuming intron nucleotide content is
272 largely selectively neutral, this may suggest a degree of expression-linked mutational-bias [76,
273 77] in this organism favoring AT mutations in introns of highly transcribed genes (or conversely,
274 elevated GC mutations at low expression levels, see below in this section). However, this
275 correlation was markedly weaker than that observed between AT3 of protein-coding genes and
276 expression across these same genes ($R=0.534$, $P<2 \times 10^{-7}$), thus providing evidence that selection
277 is a significant factor shaping AT3 [8].

278 For additional rigor in verifying the role of selection as compared to mutation in favoring
279 AT3 codons (Table 1), genes from the top 5% and lowest 5% gene expression categories were
280 placed into one of five narrow bins based on their AT-I content, specifically ≤ 0.5 , $>0.5-0.6$, $>0.6-$
281 0.7 , $>0.7-0.8$, and >0.8 . As shown in Fig. 1, for each AT-I bin, we found that AT3 of the top 5%
282 expressed genes was statistically significantly higher than that of lowly expressed genes (MWU-
283 tests P between 0.01 and <0.001). No differences in AT-I between highly and lowly expressed
284 genes were observed per bin (MWU-test $P>0.30$ in all bins, with one exception of a minimal
285 median AT-I difference of 0.019 for category 3, $P<0.05$, Fig. 1). Thus, this explicitly
286 demonstrates that within genes that have a similar background intron nucleotide composition
287 (that is, genes contained in one narrow bin of AT-I values), AT3 codons exhibit significantly
288 greater use in highly transcribed than in lowly transcribed genes. This pattern further supports
289 the interpretation that selection substantially shapes optimal codon use in *G. bimaculatus*.

290 As an additional assessment, we also considered whether the lower AT3 content of lowly
291 expressed genes (as indicated by Δ RSCU in Table 1, and in Fig. 1) could be related to biased-
292 gene conversion, which acts to enhance GC content [74, 78], in Additional File 1: Text File S1.
293 We conclude that while BGC may influence GC (and thus AT) content to some extent in this
294 taxon, it is not a major factor shaping codon use of highly versus lowly expressed genes (Table
295 1, Additional file 1: Table S2), thereby further supporting a substantive role of selection in
296 shaping AT3 optimal codon use patterns in Table 1 and Fig. 1.

297

298 ***Fop varies with tissue type and sex***

299 While the types of optimal codons identified herein were largely shared among tissues
300 (Additional file 1: Table S2), the frequency of use of these codons (Fop) varied markedly with
301 tissue type and sex in *G. bimaculatus*. In particular, Fop was markedly higher in Top5_{One-tissue}

302 genes from the testes and ovaries and the male accessory glands, than in all other six tissue types
303 (MWU-tests $P < 0.05$, Fig. 2). Thus, this suggests that genes linked to these fundamental sexual
304 structures and functions are prone to elevated optimal codon use that could, in principle, be due
305 to their essential roles in reproduction and fitness, and cost-efficient translation may be
306 particularly beneficial in the contained haploid meiotic cells [16]. Moreover, we found that the
307 Top5_{One-tissue} genes from the female somatic reproductive system had markedly higher Fop than
308 their male counterparts (MWU-test $P < 0.05$, Fig. 2). We speculate that this may reflect the
309 essential and fitness-related roles of genes involved in the insect female structures since they
310 transport and house the male sex cells and seminal fluids after mating [79, 80], possibly making
311 translational optimization more consequential to reproductive success for the female than male
312 genes. In contrast, no differences in Fop were observed with respect to sex for the brain or
313 ventral nerve cord, and the relatively low Fop values for these tissues suggest weakened selective
314 constraint on codon use of genes as compared to the gonads and to the male accessory glands
315 (MWU- tests $P < 0.05$ for the latter tissues versus the former, Fig. 2). In this regard, the data show
316 striking differences in frequency of use of the optimal codons among tissue types (Fig. 2) while
317 the identities of optimal codons themselves are largely conserved (Additional file 1: Table S2).
318 These patterns are consistent with a hypothesis that selection for translational optimization has
319 been higher for genes involved in the gonads and male accessory glands, than those from the
320 nervous system.

321 While few comparable data on multi-tissue expression and Fop are available, and
322 especially with respect to sex, a study of the male-female gonads and gonadectomized tissues in
323 *D. melanogaster* indicated that codon usage bias was lower in male than female genes [31]. This
324 pattern may be due to Hill-Robertson interference arising from adaptive evolution at linked
325 amino acid sites in the males, dragging slightly deleterious codon mutations to fixation [31].
326 However, we found an opposite pattern in the mosquito *Aedes aegypti* where optimal codon use
327 was higher in male than in female gonads [11]. Our results here, using four discrete paired male-
328 female tissue types, suggest that the only sex-related difference in Fop for *G. bimaculatus* is for
329 the somatic reproductive system (where male genes had lower Fop than female genes, Fig. 2).
330 Thus, outside the somatic reproductive system, our data show that tissue type of maximal
331 expression plays the predominant role in shaping Fop in this cricket model, rather than sex.
332 Moreover, the low Fop observed in the brain (Fig. 2) suggests that Hill-Robertson effects may be

333 greatest in this tissue type, a notion that is consistent with recent observations of a rapid rate of
334 protein sequence evolution of sex-biased brain genes in this species [64]. It is worth noting that
335 the finding that the degree of optimal codon use is particularly pronounced for genes transcribed
336 in the gonads in Fig. 2 may suggest greater absolute (but not relative) tRNA abundances of the
337 optimal codons in those reproductive tissues, which are essential for formation of the sex cells.

338

339 **Functional Roles of Optimal and Non-Optimal Codons Inferred by their Relationships to** 340 **tRNA Gene Copies**

341 The hypothesis of translational selection for efficient and/or accurate translation in an
342 organism has been thought to be substantiated by associations between optimal codon use in
343 highly expressed genes and their matching tRNA gene copy numbers in the genome [3, 5, 12, 13,
344 16, 20-25]. In some organisms however, the correspondence between optimal codon use in
345 highly expressed genes and the matching tRNA abundance has been weak [21], or not observed
346 for some (outlier) codons [81], that has been interpreted as limited support for adaptation of
347 tRNA abundance and optimal codon use [21]. However, growing evidence suggests that there is
348 a complex supply-demand relationship between codons and tRNAs that may affect multiple
349 aspects of translation [42-44, 82], such that a universal connection between optimal codons and
350 matching tRNA gene copy numbers may not always be expected [16, 42, 44]. For instance, some
351 optimal codons may obligately require wobble tRNAs (no direct matching tRNAs) [16], which
352 act to allow slow translation [48, 49], and thus a positive relationship between codon use in
353 highly expressed genes and high tRNA abundance would not be expected for those codons. In
354 turn, while non-optimal (or rare) codons may have few tRNAs, and thus act to slow translation
355 [44], in some cases they may have numerous matching tRNAs, which could conceivably allow
356 for translational upregulation of gene mRNAs using those codons [16, 45]. Given this context, to
357 allow a precise interpretation of the codon-tRNA relationships in Table 1, and given some
358 variation in terminology in the literature, we explicitly describe the codons using their Δ RSCU
359 status and their tRNA abundances as follows: Opt-codon_{↑tRNAs} are those optimal codons
360 (elevated use in highly expressed genes) that have relatively high tRNA gene copy numbers,
361 Opt-codon_{wobble}, include those optimal codons obligately requiring the use of wobble tRNAs,
362 Nonopt-codon_{↓tRNAs} are the non-optimal codons (least used in highly expressed genes) with few

363 tRNAs, and Nonopt-codon[†]tRNAs, represents non-optimal codons with abundant tRNA gene
364 copies [16].

365 To assess the relationships between the codon use and tRNA gene numbers for each
366 amino acid in Table 1, we first determined the number of tRNA genes per amino acid in the *G.*
367 *bimaculatus* genome using a recently updated version of the program tRNA-scan-SE (v. 2.0.5,
368 see Methods) [83, 84]. We report 1,391 putative tRNAs for the *G. bimaculatus* genome (Table
369 1). To evaluate the propensity for translational selection *per se*, defined as a strong relationship
370 between optimal codon use in highly expressed genes and tRNAs [5, 12, 18, 21], we compared
371 the 18 primary optimal codons to the number of tRNAs per gene. We found that for 11 of 18
372 amino acids, the primary optimal codon had the highest or near highest matching number of
373 tRNAs gene copies (≥ 18 tRNA copies) among the synonymous codons (Table 1), or Opt-
374 codon[†]tRNAs status. Thus, this concurs with a model of translational selection for accurate and/or
375 efficient translation for a majority of optimal codons in this cricket (Table 1) [5, 12, 16, 18, 21].
376 However, some optimal codons obligately required a wobble tRNA, or had Opt-codon^{wobble},
377 status, which we suggest may also serve important functional roles.

378

379 ***Some optimal codons require wobble tRNAs***

380 Seven of the 18 identified optimal codons in Table 1 had Opt-codon^{wobble} status, and had
381 no exact matching tRNAs in the genome. These included the codons AAT (Asn), GAT (Asp),
382 TGT (Cys), GGT (Gly), CAT (His), TTT (Phe), and TAT (Tyr) (Table 1). Thus, the elevated use
383 of codons with Opt-codon^{wobble} status in highly transcribed genes cannot be ascribed to
384 translational selection *per se*. We suggested in a recent report for *T. castaneum*, that optimal
385 codons obligately using wobble tRNAs may likely be employed in highly expressed genes as a
386 mechanism to slow translation, perhaps for protein folding purposes [16]. Indeed, experimental
387 research in yeast, human cells, and nematodes has shown that ribosomal translocation along the
388 mRNA is slowed by codons requiring wobble tRNAs [42, 48, 49], and thus may allow co-
389 translational protein folding. The inefficiency of wobble interactions between codons and
390 tRNAs, including chemically modified wobble tRNAs (e.g., adenosine to inosine, I34 in the
391 anticodon loop [85, 86], appears to act as a mechanism to decelerate translation as compared to
392 codons with exact tRNA matches [42, 43]. In this regard, wobble codons in highly expressed
393 genes studied here, may serve a similar function to non-optimal codons (those that have few

394 tRNAs, see below section), which growing studies suggest may regulate the rate, or rhythm, of
395 translation to allow co-translational protein folding [44, 50-53]. Notably, we found the highly
396 transcribed genes in *G. bimaculatus* were preferentially involved in protein folding as shown in
397 Table 2, and thus this comprises a primary active process within the tissues/cells under study. In
398 this regard, our collective results suggest a hypothesis that wobble codons in highly transcribed
399 genes may slow translation and effectively assist in the process of protein folding.

400 To further study the possible roles of wobble codons, we assessed the gene ontology
401 (GO) functions of the four codons with Opt-codon_{wobble} status that had the highest Δ RSCU values
402 (GGT, GAT, CAT and TAT with Δ RSCU values of +0.610, +0.520, +0.511 and +0.430
403 respectively (Table 1)) to determine if genes using these codons tended to be involved in
404 particular processes. For this, we examined the subset of highly expressed genes that were
405 especially enriched for each wobble codon (had RSCU ≥ 1.5 , where a value of 1 indicates equal
406 use of the codon per codon family, and thus ≥ 1.5 indicates a substantial elevation in use) in the
407 organism-wide dataset (Table 1), and for the genes with Top5_{One-tissue} status in the gonads
408 (Additional file 1: Table S2), which had the largest N values of any tissue type (Additional file 1:
409 Table S2; gene ontology determined from putative orthologs to *D. melanogaster* ($e < 10^{-3}$,
410 BLASTX [87]) and the program DAVID [88] and Flybase.org [89], see Methods). The results
411 are shown in Additional file 1: Table S3. The functions of the organism-wide highly expressed
412 genes with especially elevated use of the Opt-codon_{wobble} codons included ribosomal protein
413 genes, and genes involved in mitochondrion functions (Additional file 1: Table S3), thereby
414 specifically affirming that high use of these codons are apt to serve functions in these types of
415 genes. For the gonads, we found that the top GO clusters for genes with high use of GAT in the
416 ovaries (with Top5_{One-tissue} status) and of TAT in the testes (with Top5_{One-tissue} status) were
417 involved in mitosis and cell cycle functions (Additional file 1: Table S3). Thus, this pattern for
418 highly expressed gonadal genes in this cricket is in agreement with a prior experimental study
419 that suggested the use of wobble codons in genes in cultured human and yeast cells might
420 regulate the cell cycle, by controlling translation of cell-cycle genes [90]. Taken together, our
421 results are suggestive that the use of Opt-codon_{wobble} codons in highly expressed cricket genes
422 may act to slow translation as a means to regulate the level of cellular proteins, and to ensure
423 proper co-translational folding, particularly affecting genes involved in the cell-cycle
424 (Additional file 1: Table S3) and ribosomal and mitochondrial proteins (Table 2).

425

426 ***Non-optimal codons may have different functions that depend on tRNA abundance***

427 The primary non-optimal codon per amino acid was defined as the codon with the largest
428 negative Δ RSCU with a statistically significant P value [16]. With respect to the identified non-
429 optimal codons, we found striking patterns with respect to tRNAs that concur with two possible
430 functional roles, that include firstly, slowing translation, and secondly, regulating differential
431 translation of cellular mRNAs. With respect to the former case, we found two amino acids had a
432 primary non-optimal codon with Nonopt-codon \downarrow tRNAs status, that included CGC (Arg), ATC (Ile)
433 (Table 1). This suggests their infrequent use in highly expressed genes may be due to the rarity
434 or absence of matching tRNAs in the cellular tRNA pools. Moreover, these codons were not only
435 non-optimal, and thus by definition are rare in highly transcribed genes, but their exact matching
436 tRNAs were absent in the genome, and thus require wobble tRNAs, a combination that would in
437 theory make them especially prone to slowing down translation. The use of non-optimal codons
438 has been suggested to decelerate translation, which may prevent ribosomal jamming [19], and/or
439 permit proper protein folding [44, 50, 51, 91], while, as described above, the use of codons
440 requiring wobble tRNAs may also slow translation [42, 48, 49]. Thus, we propose the use of
441 these two codons in genes that have Nonopt-codon \downarrow tRNAs status, and require wobble tRNAs,
442 could play significant roles in slowing translation in highly expressed genes in *G. bimaculatus*.

443 Importantly however, the other non-optimal codons in Table 1 had tRNA counts
444 markedly higher than zero (≥ 15 gene copies; Nonopt-codon \uparrow tRNAs status). Thus, the infrequent
445 use of those non-optimal codons in the highly expressed genes is not likely to be due to a role in
446 slowing translation. In fact, the use of these codons combined with high tRNA abundance
447 suggests the potential for a high supply : demand ratio [16, 42, 45-47], a relationship that may
448 give rise to preferential translation of any highly expressed genes that contain unusually elevated
449 Nonopt-codon \uparrow tRNAs codons [16]. This proposed mechanism of up-translation using non-optimal
450 (or rare) codons has been recently suggested for stress genes in yeast [45], and for highly
451 expressed genes in the red flour beetle, wherein genes with an elevated frequency of Nonopt-
452 codon \uparrow tRNAs status codons were linked to specific biological functions [16], suggesting their
453 mRNAs may be preferentially translated. In this regard, the Nonopt-codon \uparrow tRNAs status codons in
454 *G. bimaculatus* could also have significant biological roles in up-regulation of specific cellular
455 mRNAs in this cricket model.

456 To further evaluate this possibility for *G. bimaculatus*, we studied as examples the
457 Nonopt-codon[†]tRNAs codon GTG for Val, which had an organism-wide Δ RSCU of -0.484 and 40
458 tRNAs, the codon GGC for Gly with respective values of -0.709 and 41 tRNAs (note both Val
459 and Gly are four-fold degenerate), and CTG for the six-fold degenerate Leu with a Δ RSCU of -
460 0.692 and 30 matching putative tRNAs (Table 1). These were chosen as examples due to their
461 relatively high putative tRNA counts (as compared to other Nonopt-codon[†]tRNAs codons from
462 amino acids with the same degeneracy level). For each of these codons, we examined those
463 Top5^{One tissue} genes (only in the top 5% expression in one tissue type) in the gonads that had
464 RSCU value ≥ 1.5 , indicating enhanced use. The results are shown in Table 3. We found that
465 genes preferentially using Nonopt-codon[†]tRNA codons were associated with a diverse range of
466 functions. For example, for the ovaries, the highly expressed genes that preferentially used the
467 Nonopt-codon[†]tRNAs codon GTG (for Val) included a match to *Bicaudal C* (*BicC*), which is
468 involved in oogenesis [92]. Remarkably, this ovary gene also had elevated use of the wobble
469 codons GGC and CTG (Table 1). Similarly, for the ovaries, an ortholog of *santa-maria*, which has
470 been associated with phototransduction [93] and apoptosis [94], had elevated use of each of the
471 wobble codons GTG, GGC and CTG. The fact that both *BicC* and *santa-maria* each have high
472 use of all three of these Nonopt-codon[†]tRNAs codons, which by definition have abundant matching
473 tRNA genes, suggests their gene transcripts are preferentially translated in the ovary as compared
474 to other transcripts in the transcript pool. For CTG (Leu), the Top5^{One-tissue} genes in the ovaries
475 preferentially using this codon with Nonopt-codon[†]tRNAs status included another apoptosis gene,
476 *apoptosis inducing factor* (*AIF*) [95], which also had elevated use of GGC for Gly, suggesting
477 these codons may facilitate apoptosis in the female gonad cells. With respect to the testis, GTG
478 (Val) was preferentially used in genes such as *belle*, which is involved in male germ-line stem
479 cell development [96, 97] and *no child left behind* (*nclb*), involved in male gonad development
480 [98], suggesting that use of this non-optimal codon may promote translation of these particular
481 transcripts in the male gonadal mRNA pools. Enhanced use of GGC and CTG in testes genes
482 matching *Dual-specificity tyrosine phosphorylation-regulated kinase 2* (*Dyrk2*), which is
483 involved in apoptosis and sensory roles [99, 100], and *short spindle 3* (*ssp3*), involved in male
484 meiosis [101] (Table 3), infers that these two codons may promote translation of apoptosis and
485 meiotic proteins in the testes. When taken together, these patterns in *G. bimaculatus*, similar to
486 recent findings in *T. castaneum* [16], suggest that the combination of elevated use of non-optimal

487 codons and a high supply of tRNAs may plausibly be involved in preferential translation of the
488 transcripts of specific genes in this system, particularly for apoptosis genes and genes with
489 female and male gonadal functions (Table 3).

490

491 **Amino Acid Use, Biosynthesis Costs, and tRNA Gene Copies have Interdependently** 492 **Evolved**

493 Next, we asked whether amino acid use in the highly expressed genes in *G. bimaculatus*
494 (top 5% using the organism-wide assessment) varied with their size/complexity (S/C) scores,
495 which were developed to quantify the relative biosynthesis costs of different amino acids [56],
496 hydrophathy, or with their broad role in protein folding properties [102, 103] (Additional file 1:
497 Table S4). As shown in Fig. 3, for highly expressed genes the amino acid usage (across all 20
498 amino acids) was not correlated to hydrophathy (Spearman's correlation across all 777 organism-
499 wide highly expressed genes $P > 0.60$) and showed no broad relationship to specific protein
500 folding properties (ranked ANOVA $P > 0.05$ between groups, Fig. 3BC). However, a very strong
501 negative correlation was observed between amino acid use and S/C scores across the 20 amino
502 acids (Spearman's $R = -0.87$, $P < 2 \times 10^{-7}$, Fig. 3A, Table 4; see also [10]). An inverse relationship
503 between S/C score and the frequency of the 20 amino acids was also observed across all 15,539
504 studied *G. bimaculatus* genes irrespective of expression level (for all genes $R = -0.70$, $P = 4 \times 10^{-4}$,
505 Additional file 1: Fig. S1), but the correlation was stronger in the subset of highly expressed
506 genes, suggesting that the connection between amino acid use and S/C scores is ameliorated with
507 elevated transcription. Thus, these patterns both at the genome-wide level and using highly
508 expressed genes measured across nine tissue types, indicate preferential use of low-cost amino
509 acids in genes producing abundant mRNAs.

510 To further decipher this relationship, we compared amino acid usage using the organism-
511 wide highest and lowest expressed genes (top and lowest 5%, averaged across nine tissues). As
512 shown in Table 4, we found that 19 of 20 amino acids had a statistically different frequency
513 between the most and least transcribed genes in the genome (t-tests $P < 0.05$), with the only
514 exception being Thr. The amino acids with the largest increase in frequency in highly expressed
515 genes (as compared to lowly expressed) were Ile (S/C score=16.04; with 49.0% greater use under
516 high expression) and Lys (30.14; 49.1% greater use under high expression), suggesting that
517 enhanced use of these amino acids with intermediate S/C scores may be more crucial to efficient

518 translation or function of abundant transcripts, than the use of those with the lowest possible S/C
519 scores in this taxon. We note this is consistent with an earlier analysis based on a partial
520 transcriptome from one pooled ovary/embryo sample and without tRNA data in that study, where
521 amino acids with intermediate S/C scores Glu, Asp, and Asn were preferred [10], that all had
522 >22% increased use under high transcription here. This type of complex relationship between
523 S/C score and amino acid use has also been suggested in spiders [55].

524 Under a null hypothesis of equal usage of each of 20 amino acids, we would assume a
525 frequency of 5% for every amino acid per gene, with values above and below this threshold
526 indicating favored and disfavored usage respectively. In this context, we observed that for the
527 five highest cost amino acids (Tyr, Cys, His, Met and Trp, S/C scores of 57.00 to 73.00), the
528 average usage was less than 5% (between 1.18 and 3.10%) in both the highly and lowly
529 expressed genes (Table 4), indicating these biochemically costly amino acids are consistently
530 rarely used in this taxon. Taken together, organism-wide highly expressed genes in *G.*
531 *bimaculatus* exhibit a pattern of elevated use of amino acids with low S/C scores (Fig. 3A), and
532 also exhibit elevated use of specific amino acids with intermediate S/C scores (Table 4), and very
533 low use of the highest cost amino acids. We speculate that the pattern of favored use of some
534 intermediate cost amino acids may be due to the roles of these amino acids in protein folding
535 (e.g., beta and alpha folding respectively, Additional file 1: Table S4) and thus their use may
536 ensure proper function of abundantly produced gene products.

537 With respect to tRNA abundances, we found that amino acid frequencies in Table 4 were
538 positively correlated to the tRNA gene counts per amino acid (the tRNA counts included all
539 those matching any of synonymous codons per amino acid) in *G. bimaculatus*. The correlation
540 was observed both for the highly and for the lowly expressed genes (Spearman's Ranked R=0.65
541 and 0.75, $P < 0.05$, Table 4). Thus, this suggests the frequency of amino acid use within genes is
542 connected to its tRNA abundance in this organism. However, despite being correlated in both
543 groups (high and low expressed genes) in this cricket species, we suggest that the relationship is
544 apt to be most beneficial to the organism by reducing the translational costs of genes that are
545 highly transcribed, as these genes should presumably be most commonly translated.

546 We next asked whether tRNA abundance, or gene copy number, was connected to S/C
547 scores in *G. bimaculatus*. Indeed, the 20 amino acids showed a striking tendency to be inversely
548 connected to the total tRNA counts per amino acid in the organism-wide highly expressed genes

549 (Spearman's $R=-0.52$, $P=0.02$, Fig. 4). Thus, the abundance of tRNAs in the genome is directly
550 connected to how biochemically costly an amino acid is to produce by the organism. While
551 comparable studies of relationships between biosynthetic amino acid costs and tRNAs are
552 uncommon, a similar negative pattern has been observed in a study from beetles [21], suggesting
553 this phenomenon may be shared among diverse insects. Taking all our results in combination, it
554 is evident that amino acid frequency is positively correlated to the matching tRNA gene counts
555 (Table 4), and negatively correlated to S/C scores (Fig. 3A, Additional file 1: Fig. S1), and that
556 tRNA gene counts per amino acid are negatively related to S/C scores (Fig. 4). In other words,
557 genes exhibit a tendency for preferred use of low cost amino acids that have abundant tRNAs.
558 We therefore suggest the hypothesis that all three parameters, amino acid frequency, tRNA genes
559 in the genome, and biochemical costs, have evolved interdependently for translational
560 optimization in *G. bimaculatus*.

561 It should be noted that while we specify herein that our tRNAs counts obtained from
562 tRNA-scan-SE (v. 2.0.5) [83, 84] from the recently available cricket genome [65] are considered
563 preliminary predictions in this study (see Methods, Table 1), the accuracy of this list is
564 substantiated by the marked correlation of tRNA gene counts with S/C scores (Fig. 4) and with
565 amino acid frequency (Table 4). In this regard, we consider the relative tRNA counts apt to
566 provide an appropriate and accurate profile for *G. bimaculatus*.

567

568 ***Variation in amino acid use with respect to sex and tissue type***

569 Finally, we determined whether amino acid frequency per gene varied among tissue type
570 or sex for those genes with Top5One-tissue status. The results for amino acid frequency are shown
571 in Additional file 1: Table S5, and correlations between use for each sex per tissue type are
572 provided in Additional file 1: Table S6. For each sex, we found strong correlations in the
573 frequency of amino acid use (across 20 amino acids) for all paired contrasts of tissues, with
574 Spearman R values between 0.861 and 0.98 ($P<2\times 10^{-6}$). This suggests the relative amino acid
575 use is largely consistent among highly expressed genes from all tissue types. However, the R
576 values were weakest ($R<0.9$) for contrasts of the male gonad to all other tissues, suggesting a
577 possible testis-effect on amino acid use. In terms of differences between sexes, we determined
578 the percent difference in frequency of amino acid use between females and males for each tissue
579 type (Additional file 1: Table S5). We found that amino acid use varied between the sexes, with

580 between two to six amino acids per tissue type (gonad, somatic reproductive system, brain,
581 ventral nerve cord) exhibiting statistically significant differences between sexes. As an example,
582 for the Top5_{One-tissue} genes from the brain which had six amino acids with statistically significant
583 differences between males and females, we found that some amino acids, namely Arg and Tyr,
584 had in excess of 21% difference in their use between the sexes in *G. bimaculatus* (t-test $P < 0.05$;
585 Additional file 1: Table S5), thus revealing particularly marked variation for this tissue. In this
586 regard, there are non-negligible differences in amino acid use between the sexes, particularly for
587 the brain, suggesting that high expression in a particular sex may be a significant factor
588 contributing to amino acid use.

589

590 **Conclusions**

591 Our collective results herein strongly suggest a model whereby codon use and amino acid
592 use have adapted to facilitate multiple functions in highly expressed genes the cricket *G.*
593 *bimaculatus*. Specifically, we showed that optimal codons are largely shared across diverse
594 tissue types and both sexes in this organism, and are likely shaped by selective pressures (Table
595 1, Fig. 1, Additional file 1: Table S2). Further, we revealed that a majority of optimal codons
596 have abundant tRNA gene copies (Table 1), which is concordant with functional roles in
597 translational optimization [12, 22]. Importantly however, we found that a substantial subset of
598 optimal codons obligately require wobble tRNAs (Table 1), suggesting their use may have
599 evolved as a mechanism to slow translation in highly transcribed genes, a notion that is
600 supported by available experimental *in vivo* translation research [42, 48, 49]. These wobble
601 codons may facilitate protein folding and/or be involved in regulation of genes such as cell-cycle
602 genes (Additional file 1: Table S3). In turn, we demonstrated that non-optimal codons,
603 particularly those that have few or no matching tRNAs gene copies in the genome (Table 1), may
604 also act to slow translation, concurring with the notion that non-optimal codons may limit
605 ribosome jamming or protein folding [19, 44, 50-53]. Crucially however, our data revealed that
606 not all non-optimal codons are apt to have this putative function. Rather, we find that many non-
607 optimal codons have abundant directly matching tRNA genes (Table 1), and are linked to
608 specific types of gene ontology functions such as apoptosis and gonadal functions (Table 3).
609 Thus, the non-optimal codons with abundant tRNAs likely provide a major organismal
610 mechanism to promote the upregulation of specific mRNAs in the cellular mRNA pool, agreeing

611 with a model proposed in some recent studies [16, 45]. Finally with respect to amino acid use,
612 our data suggest a hypothesis that amino acid use, biochemical costs of amino acids [56], and
613 tRNA gene counts in the genome (Fig. 3A, Fig. 4, Table 4) have interdependently evolved as a
614 mechanism for translational optimization of highly expressed genes in *G. bimaculatus*.

615 Future research should include the direct quantification of tRNAs in different tissue
616 types, a method that remains under development and debate [33, 45, 68, 104], to assess whether
617 those results add support to the conclusion of similar relative tRNA abundances among amino
618 acids across tissue types and sexes in this cricket. Moreover, further studies should be conducted
619 of the frequencies of optimal, as well as non-optimal, codons and their relationships to tRNA
620 abundances and gene functionalities, in a wider range of multicellular organisms. Such research
621 will reveal whether the phenomena observed herein are shared across divergent systems.

622

623

624 **Materials and Methods**

625 **Biological Samples and RNA-seq**

626 Gene expression level was determined for all 15,539 *G. bimaculatus* protein-coding
627 genes (CDS, longest CDS per gene) [65] that had a start codon and were >150bp. RNA-seq was
628 obtained for four adult male and female tissue types, the gonad (testis for males, ovaries for
629 females), somatic reproductive system, brain and ventral nerve cord and for the male accessory
630 glands (Additional file 1: Table S1) as described previously [64]. The expression level of each *G.*
631 *bimaculatus* gene was determined by mapping reads per RNA-seq dataset per tissue to the
632 complete CDS list using Geneious Read Mapper [105], to determine FPKM per gene. FPKM
633 was robust to mapping programs, and other common mappers including BBmap
634 (<https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/>) and Bowtie2 [106]
635 yielded similar results [64].

636 Optimal codons for the organism-wide analysis (averaged expression across all nine
637 tissues) and Δ RSCU is described in the “Results and Discussion”. For each codon, using
638 Δ RSCU = RSCU_{Mean Highly Expressed CDS} - RSCU_{Mean Low Expressed CDS}, t-tests were conducted between
639 highly and lowly expressed genes to assess statistical significance. To isolate the effect of each
640 individual tissue type, the optimal codons were determined separately for each of the nine tissues
641 under study (males and females for each tissue type, and male accessory glands). It has been
642 suggested that optimal codon use in a gene largely depends on the tissue type in which it is
643 maximally transcribed [16, 30]. Accordingly, to identify optimal codons for each tissue type, we
644 examined those genes that were in the top 5% expression in that one tissue type and not in the
645 top 5% expression for any of the remaining eight tissues (denoted as Top5_{One-tissue}) versus those
646 with the lowest 5% expression (or all those tied with the FPKM cutoff of the lowest 5% [16]).
647 Using these highly and lowly expressed genes per tissue, the Δ RSCU was determined as
648 described for the organism-wide optimal codons.

649 The frequency of optimal codons (Fop) [4] for each gene under study was determined,
650 using the identified optimal codons, in the program CodonW (Peden 1999). Fop was then
651 compared for genes with high transcription in the various tissue types and two sexes in *G.*
652 *bimaculatus*.

653

654 **Intron Analysis**

655 We compared the AT (or GC) content of introns, which are thought to largely reflect the
656 innate mutational pressures on the nucleotide content of genes [74, 107, 108], to the AT3
657 content (third nucleotide position) of CDS of highly and lowly expressed genes for the *G.*
658 *bimaculatus* organism-wide optimal codons [16]. For this, using the genomic data for *G.*
659 *bimaculatus*, we extracted the introns for all genes (with introns), and retained those >50bp after
660 trimming of 10bp from the 5' and 3' ends which may contain regulatory/conserved regions [74]
661 (and studied the longest intron per gene). For additional stringency, given that highly transcribed
662 genes have been suggested to exhibit mutational biases (e.g., C to T) within a small number of
663 organisms (e.g., *E. coli*, humans [76, 77]), we tested whether there was a correlation between
664 gene expression and intron AT content in *G. bimaculatus*. To further assess the role of selection,
665 as compared to mutation, in favoring AT3 codons (Table 1), genes from the top 5% and lowest
666 5% gene expression categories were placed into one of five bins based on their AT-I content as
667 shown in Fig. 1.

668

669 **tRNA Gene Copies**

670 The number of tRNA genes per amino acid in the *G. bimaculatus* genome was
671 determined using the recently updated version of tRNA-scan-SE (v. 2.0.5) [83, 84]. The
672 Eukaryotic filter called EukHighConfidenceFilter was used, which was designed to narrow the
673 tRNA-scan output to a conservative high confidence tRNA [83] (used at default settings with the
674 exception of ml -1). We note that since the rigor of the updated program has not been explicitly
675 tested in insects outside *Drosophila* (P. Chan, personal communication), we consider the tRNA
676 predictions preliminary, and focus on the relative values of tRNAs among codons and amino
677 acids. The accuracy of the predictions, however, is strongly supported by the correlations
678 between tRNA gene copy numbers and amino acid costs and amino acid frequency (see section
679 “*Amino Acid Use, Biosynthesis Costs, and tRNA Gene Copies have Interdependently Evolved*”).
680 The filter acted to reduced the absolute counts of tRNAs per amino acid to the high confidence
681 dataset. Nonetheless, the tRNA counts with and without the filter were strongly correlated across
682 amino acids (Spearman’s Ranked R =0.90, P<2X10⁻⁷), and thus relative gene counts remain
683 intact using both measures.

684

685 **Amino Acid Use**

686 The frequency of each of the 20 amino acids in protein-coding genes in an organism may
687 be influenced by factors such as their size/complexity Dufton scores (which range from 1 to 73
688 depending on the amino acid, [56]), as well as hydropathy (where positive hydrophobicity values
689 indicate hydrophobic nature, while negative suggest a hydrophilic amino acid [102, 103]), and/or
690 their role in protein folding structures (alpha helices, beta sheets, or breakers used to affect
691 bonding in helices) [103]. We thus aimed to study each of these parameters, using established
692 values per amino acid shown in Table S4. We evaluated whether amino acid frequency in
693 proteins of highly transcribed genes at an organism-wide level in *G. bimaculatus* (top 5%
694 average expression across all eight male and female tissue) was correlated to S/C score [56], as
695 well as hydropathy and protein folding characteristics [56, 102, 103]. In addition, we assessed
696 and compared amino acid use per tissue type/sex by examining genes with Top5One-tissue status per
697 tissue type.

698

699 **Gene Ontology**

700 For gene ontology functions, we used the gene ontology from the fly *D. melanogaster*,
701 which comprises the most well studied insect genome to date [89]. For this, we conducted a
702 BLAST search of the full *G. bimaculatus* CDS list under study to *D. melanogaster* CDS list
703 (version 6.29 [89]) using BLASTX [87], applying a cutoff of $e < 10^{-3}$. For those genes having
704 matches within these criteria, the *D. melanogaster* gene identifiers of were then input into the
705 program DAVID [88] for gene ontology analyses and searched in FlyBase [89].

706 **List of abbreviations**

707 Top5One-tissue, genes with an expression level in the top 5% in one tissue type only, and not in the
708 other eight tissues

709 FPKM, frequency per kilobase million

710 MWU-test, Mann-Whitney U-test

711

712 **Declarations**

713 *Ethics approval and consent to participate*

714 Not applicable.

715 *Consent for publication*

716 Not applicable.

717 *Availability of data and material*

718 All RNA-seq data under study are described in Additional file 1: Table S1 and are available at
719 the Short Read Archive (SRA) under the project identifier PRJNA564136.

720 *Competing interests*

721 The authors declare they have no competing interests.

722 *Funding*

723 This work was supported by funds from Harvard University.

724 *Acknowledgements*

725 The authors thank Dr. Guillem Ylla for providing early access to the assembled *G. bimaculatus*
726 genome and members of the Extavour lab for discussions. The services of the Bauer core
727 sequencing facility at Harvard University are appreciated.

728 *Authors' contributions*

729 CAW, AK and CGE designed the study. AK reared *G. bimaculatus* and sampled tissues for
730 RNA-seq. CAW analyzed the data and wrote the manuscript with contributions by AK, NC and
731 CGE. NC contributed to GO analysis. All authors read and approved the final manuscript.

732

733 **Additional Files**

734 Additional File 1: The file contains the Supplementary Tables, Figures and Text which are
735 denoted and Tables S1 to S6, Figure S1, and Text File S1.

References

1. Plotkin JB, Kudla G: **Synonymous but not the same: the causes and consequences of codon bias.** *Nature Reviews Genetics* 2011, **12**(1):32-42.
2. Whittle CA, Sun Y, Johannesson H: **Evolution of synonymous codon usage in *Neurospora tetrasperma* and *Neurospora discreta*.** *Genome Biology and Evolution* 2011, **3**:332-343.
3. Percudani R, Pavesi A, Ottonello S: **Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*.** *Journal of Molecular Biology* 1997, **268**(2):322-330.
4. Ikemura T: **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.** *Journal of Molecular Biology* 1981, **151**(3):389-409.
5. Akashi H: **Gene expression and molecular evolution.** *Current Opinion in Genetics & Development* 2001, **11**:660-666.
6. Satapathy SS, Powdel BR, Buragohain AK, Ray SK: **Discrepancy among the synonymous codons with respect to their selection as optimal codon in bacteria.** *DNA Research* 2016, **23**:441-449.
7. Ingvarsson PK: **Molecular evolution of synonymous codon usage in *Populus*.** *BMC Evol Biol* 2008, **8**:307.
8. Qiu S, Bergero R, Zeng K, Charlesworth D: **Patterns of codon usage bias in *Silene latifolia*.** *Molecular Biology and Evolution* 2011, **28**(1):771-780.
9. Cutter AD, Wasmuth JD, Blaxter ML: **The evolution of biased codon and amino acid usage in nematode genomes.** *Molecular Biology and Evolution* 2006, **23**(12):2303-2315.
10. Whittle CA, Extavour CG: **Codon and amino acid usage are shaped by selection across divergent model organisms of the Pancrustacea.** *G3: Genes, Genomes, Genetics* 2015, **5**(11):2307-2321.
11. Whittle CA, Extavour CG: **Rapid Evolution of Ovarian-Biased Genes in the Yellow Fever Mosquito (*Aedes aegypti*).** *Genetics* 2017, **206**(4):2119-2137.
12. Duret L: **tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes.** *Trends in Genetics* 2000, **16**(7):287-289.
13. Behura SK, Severson DW: **Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in *Aedes aegypti* and *Anopheles gambiae*.** *Insect Molecular Biology* 2011, **20**:177-187.
14. Whittle CA, Malik MR, Krochko JE: **Gender-specific selection on codon usage in plant genomes.** *BMC Genomics* 2007, **8**:169-179.
15. Duret L, Mouchiroud D: **Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*.** *Proc Natl Acad Sci U S A* 1999, **96**(8):4482-4487.
16. Whittle CA, Kulkarni A, Extavour CG: **Evidence of multifaceted functions of codon usage in translation within the model beetle *Tribolium castaneum*.** *DNA Research* 2019, **26**(6):473-484.
17. Du MZ, Wei W, Qin L, Liu S, Zhang AY, Zhang Y, Zhou H, Guo FB: **Co-adaptation of tRNA gene copy number and amino acid usage influences translation rates in three life domains.** *DNA Research* 2017, **24**(6):623-633.

18. Sharp PM, Tuohy TM, Mosurski KR: **Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes.** *Nucleic Acids Research* 1986, **14**(13):5125-5143.
19. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y: **An evolutionarily conserved mechanism for controlling the efficiency of protein translation.** *Cell* 2010, **141**(2):344-354.
20. Cognat V, Deragon JM, Vinogradova E, Salinas T, Rémacle C, Marechal-Drouard L: **On the evolution and expression of *Chlamydomonas reinhardtii* nucleus-encoded transfer RNA genes.** *Genetics* 2008, **179**(1):113-123.
21. Williford A, Demuth JP: **Gene expression levels are correlated with synonymous codon usage, amino acid composition, and gene architecture in the red flour beetle, *Tribolium castaneum*.** *Molecular Biology and Evolution* 2012, **29**(12):3755-3766.
22. Ikemura T: **Codon usage and tRNA content in unicellular and multicellular organisms.** *Molecular Biology and Evolution* 1985, **2**(1):13-34.
23. Rocha EP: **Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization.** *Genome Research* 2004, **14**(11):2279-2286.
24. Moriyama EN, Powell JR: **Codon usage bias and tRNA abundance in *Drosophila*.** *Journal of Molecular Evolution* 1997, **45**(5):514-523.
25. Powell JR, Moriyama EN: **Evolution of codon usage bias in *Drosophila*.** *Proc Natl Acad Sci U S A* 1997, **94**(15):7784-7790.
26. Ellegren H, Parsch J: **The evolution of sex-biased genes and sex-biased gene expression.** *Nature Reviews Genetics* 2007, **8**(9):689-698.
27. Ingleby FC, Flis I, Morrow EH: **Sex-biased gene expression and sexual conflict throughout development.** *Cold Spring Harbor Perspectives in Biology* 2014, **7**(1):a017632.
28. Grath S, Parsch J: **Sex-Biased Gene Expression.** *Annual Review of Genetics* 2016, **50**:29-44.
29. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S: **Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees.** *Science* 2005, **309**:1850-1854.
30. Camiolo S, Farina L, Porceddu A: **The relation of codon bias to tissue-specific gene expression in *Arabidopsis thaliana*.** *Genetics* 2012, **192**(2):641-649.
31. Hambuch TM, Parsch J: **Patterns of synonymous codon usage in *Drosophila melanogaster* genes with sex-biased expression.** *Genetics* 2005, **170**(4):1691-1700.
32. Payne BL, Alvarez-Ponce D: **Codon usage differences among genes expressed in different tissues of *Drosophila melanogaster*.** *Genome Biology and Evolution* 2019(11):1054-1065.
33. Dittmar KA, Goodenbour JM, Pan T: **Tissue-specific differences in human transfer RNA expression.** *PLoS Genetics* 2006, **2**(12):e221.
34. Plotkin JB, Robins H, Levine AJ: **Tissue-specific codon usage and the expression of human genes.** *Proc Natl Acad Sci U S A* 2004, **101**(34):12588-12591.
35. Liu Q: **Mutational bias and translational selection shaping the codon usage pattern of tissue-specific genes in rice.** *PLoS One* 2012, **7**(10):e48295.

36. Matsumoto Y, Sakai M: **Brain control of mating behavior in the male cricket *Gryllus bimaculatus* DeGeer: brain neurons responsible for inhibition of copulation actions.** *Journal of Insect Physiology* 2000, **46**(4):539-552.
37. Sakai M, Kumashiro M, Matsumoto Y, Ureshi M, Otsubo T: **Reproductive Behavior and Physiology in the Cricket *Gryllus bimaculatus*.** In: *The Cricket as a Model Organism: Development, Regeneration and Behavior*. Edited by Horch HW, Mito T, Popadic A, Ohuchi H, Noji S, vol. : Springer; 2017: 245-269.
38. Haberkern H, Hedwig B: **Behavioural integration of auditory and antennal stimulation during phonotaxis in the field cricket *Gryllus bimaculatus*.** *Journal of Experimental Biology* 2016, **219**(Pt 22):3575-3586.
39. Wang B, Shao ZQ, Xu Y, Liu J, Liu Y, Hang YY, Chen JQ: **Optimal codon identities in bacteria: implications from the conflicting results of two different methods.** *PLoS One* 2011, **6**(7):e22714.
40. Hershberg R, Petrov DA: **Selection on codon bias.** *Annual Review of Genetics* 2008, **42**:287-299.
41. Hershberg R, Petrov DA: **General rules for optimal codon choice.** *PLoS Genetics* 2009, **5**(7):e1000556.
42. Stein KC, Frydman J: **The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis.** *Journal of Biological Chemistry* 2019, **294**(6):2076-2084.
43. Brule CE, Grayhack EJ: **Synonymous Codons: Choose Wisely for Expression.** *Trends in Genetics* 2017, **33**(4):283-297.
44. Quax T, Claassens N, Soll D, van der Oost J: **Codon Bias as a Means to Fine-Tune Gene Expression.** *Molecular Cell* 2015, **59**:149-161.
45. Torrent M, Chalancon G, de Groot NS, Wuster A, Madan Babu M: **Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions.** *Sci Signal* 2018, **11**(546):DOI: 10.1126/scisignal.aat6409.
46. Gingold H, Dahan O, Pilpel Y: **Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome.** *Nucleic Acids Research* 2012, **40**(20):10053-10063.
47. Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF: **Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression.** *Cell* 2016, **165**(6):1416-1427.
48. Stadler M, Fire A: **Wobble base-pairing slows in vivo translation elongation in metazoans.** *RNA* 2011, **17**(12):2063-2073.
49. Letzring DP, Dean KM, Grayhack EJ: **Control of translation efficiency in yeast by codon-anticodon interactions.** *RNA* 2010, **16**(12):2516-2528.
50. Zalucki YM, Jennings MP: **Experimental confirmation of a key role for non-optimal codons in protein export.** *Biochemical and Biophysical Research Communications* 2007, **355**(1):143-148.
51. Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, Liu Y: **Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding.** *Molecular Cell* 2015, **59**(5):744-754.
52. Pechmann S, Frydman J: **Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding.** *Nature Structural and Molecular Biology* 2013, **20**(2):237-243.

53. Zhou M, Wang T, Fu J, Xiao G, Liu Y: **Nonoptimal codon usage influences protein structure in intrinsically disordered regions.** *Molecular Microbiology* 2015, **97**(5):974-987.
54. Whittle CA, Extavour CG: **Codon and Amino Acid Usage Are Shaped by Selection Across Divergent Model Organisms of the Pancrustacea.** *G3 (Bethesda)* 2015, **5**(11):2307-2321.
55. Whittle CA, Extavour CG: **Expression-Linked Patterns of Codon Usage, Amino Acid Frequency, and Protein Length in the Basally Branching Arthropod *Parasteatoda tepidariorum*.** *Genome Biology and Evolution* 2016, **8**(9):2722-2736.
56. Dufton MJ: **Genetic code synonym quotas and amino acid complexity: cutting the cost of proteins?** *Journal of Theoretical Biology* 1997, **187**(2):165-173.
57. Gaunt MW, Miles MA: **An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks.** *Molecular Biology and Evolution* 2002, **19**(5):748-761.
58. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG *et al*: **Phylogenomics resolves the timing and pattern of insect evolution.** *Science* 2014, **346**(6210):763-767.
59. Zeng V, Ewen-Campen B, Horch HW, Roth S, Mito T, Extavour CG: **Developmental gene discovery in a hemimetabolous insect: de novo assembly and annotation of a transcriptome for the cricket *Gryllus bimaculatus*.** *PLoS One* 2013, **8**(5):e61479.
60. Fisher HP, Pascual MG, Jimenez SI, Michaelson DA, Joncas CT, Quenzer ED, Christie AE, Horch HW: **De novo assembly of a transcriptome for the cricket *Gryllus bimaculatus* prothoracic ganglion: An invertebrate model for investigating adult central nervous system compensatory plasticity.** *PLoS One* 2018, **13**(7):e0199070.
61. Mito T, Noji S: **The Two-Spotted Cricket *Gryllus bimaculatus*: An Emerging Model for Developmental and Regeneration Studies.** *Cold Spring Harbor Protocols* 2008, **2008**:pdb emo110.
62. Donoughe S, Extavour CG: **Embryonic development of the cricket *Gryllus bimaculatus*.** *Developmental Biology* 2016, **411**(1):140-156.
63. Nakamura T, Extavour CG: **The transcriptional repressor Blimp-1 acts downstream of BMP signaling to generate primordial germ cells in the cricket *Gryllus bimaculatus*.** *Development* 2016, **143**(2):255-263.
64. Whittle CA, Kulkarni A, Extavour CG: **Sex-biased genes expressed in the cricket brain evolve rapidly.** *BioRxiv* 2020, www.biorxiv.org/content/10.1101/2020.07.07.192039v1
65. Ylla G, Nakamura T, Itoh T, Kajitani R, Toyoda A, Tomonari S, Bando T, Ishimaru Y, Watanabe T, Fuketa M *et al*: **Cricket genomes: the genomes of future food.** *bioRxiv* 2020:2020.2007.2007.191841.
66. Shields DC, Sharp PM, Higgins DG, Wright F: **"Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons.** *Molecular Biology and Evolution* 1988, **5**(6):704-716.
67. Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE: **Variation in the strength of selected codon usage bias among bacteria.** *Nucleic Acids Research* 2005, **33**(4):1141-1153.
68. Whittle CA, Kulkarni A, Extavour CG: **Absence of a faster-X effect in beetles (*Tribolium*, Coleoptera).** *G3: Genes, Genomes, Genetics* 2020, **10**:1125–1136.

69. Cutter AD: **Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate.** *Molecular Biology and Evolution* 2008, **25**(4):778-786.
70. Haddrill PR, Charlesworth B, Halligan DL, Andolfatto P: **Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content.** *Genome Biology* 2005, **6**(8):R67.
71. D'Onofrio G, Ghosh TC, Saccone S: **Different functional classes of genes are characterized by different compositional properties.** *FEBS Letters* 2007, **581**(30):5819-5824.
72. Behura SK, Singh BK, Severson DW: **Antagonistic relationships between intron content and codon usage bias of genes in three mosquito species: functional and evolutionary implications.** *Evolutionary Applications* 2013, **6**(7):1079-1089.
73. Zeng K, Charlesworth B: **Studying patterns of recent evolution at synonymous sites and intronic sites in *Drosophila melanogaster*.** *Journal of Molecular Evolution* 2010, **70**(1):116-128.
74. Chamary JV, Hurst LD: **Similar rates but different modes of sequence evolution in introns and at exonic silent sites in rodents: evidence for selectively driven codon usage.** *Molecular Biology and Evolution* 2004, **21**(6):1014-1023.
75. Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA: **Selection for short introns in highly expressed genes.** *Nature Genetics* 2002, **31**(4):415-418.
76. Mugal CF, von Grunberg HH, Peifer M: **Transcription-induced mutational strand bias and its effect on substitution rates in human genes.** *Molecular Biology and Evolution* 2009, **26**(1):131-142.
77. Beletskii A, Bhagwat AS: **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* 1996, **93**(24):13919-13924.
78. Pouyet F, Mouchiroud D, Duret L, Semon M: **Recombination, meiotic expression and human codon usage.** *eLife* 2017, **6**.
79. Degner EC, Harrington LC: **A mosquito sperm's journey from male ejaculate to egg: Mechanisms, molecules, and methods for exploration.** *Molecular Reproduction and Development* 2016, **83**(10):897-911.
80. Pascini TV, Martins GF: **The insect spermatheca: an overview.** *Zoology* 2017, **121**:56-71.
81. Wright SI, Yau CB, Looseley M, Meyers BC: **Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*.** *Molecular Biology and Evolution* 2004, **21**(9):1719-1726.
82. Guimaraes JC, Mittal N, Gnann A, Jedlinski D, Riba A, Buczak K, Schmidt A, Zavolan M: **A rare codon-based translational program of cell proliferation.** *Genome Biology* 2020, **21**(1):44.
83. Chan PP, Lowe TM: **tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences.** *Methods in Molecular Biology* 2019, **1962**:1-14.
84. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Research* 1997, **25**(5):955-964.
85. Torres AG, Pineyro D, Filonava L, Stracker TH, Battlle E, Ribas de Pouplana L: **A-to-I editing on tRNAs: biochemical, biological and evolutionary implications.** *FEBS Lett* 2014, **588**(23):4279-4286.

86. Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L: **A role for tRNA modifications in genome structure and codon usage.** *Cell* 2012, **149**(1):202-213.
87. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Research* 1997, **25**(17):3389-3402.
88. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44-57.
89. Gramates LS, Marygold SJ, Santos GD, Urbano JM, Antonazzo G, Matthews BB, Rey AJ, Tabone CJ, Crosby MA, Emmert DB *et al*: **FlyBase at 25: looking to the future.** *Nucleic Acids Research* 2017, **45**:D663-D671.
90. Frenkel-Morgenstern M, Danon T, Christian T, Igarashi T, Cohen L, Hou YM, Jensen LJ: **Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels.** *Molecular Systems Biology* 2012, **8**:572.
91. Zhao F, Yu CH, Liu Y: **Codon usage regulates protein structure and function by affecting translation elongation speed in *Drosophila* cells.** *Nucleic Acids Research* 2017, **45**(14):8484-8492.
92. Saffman EE, Styhler S, Rother K, Li W, Richard S, Lasko P: **Premature translation of *oskar* in oocytes lacking the RNA-binding protein bicaudal-C.** *Molecular and Cellular Biology* 1998, **18**(8):4855-4862.
93. Wang T, Jiao Y, Montell C: **Dissection of the pathway required for generation of vitamin A and for *Drosophila* phototransduction.** *Journal of Cell Biology* 2007, **177**(2):305-316.
94. Herboso L, Talamillo A, Perez C, Barrio R: **Expression of the Scavenger Receptor Class B type I (SR-BI) family in *Drosophila melanogaster*.** *International Journal of Developmental Biology* 2011, **55**(6):603-611.
95. Stambolsky P, Weisz L, Shats I, Klein Y, Goldfinger N, Oren M, Rotter V: **Regulation of AIF expression by p53.** *Cell Death and Differentiation* 2006, **13**(12):2140-2149.
96. Johnstone O, Deuring R, Bock R, Linder P, Fuller MT, Lasko P: **Belle is a *Drosophila* DEAD-box protein required for viability and in the germ line.** *Developmental Biology* 2005, **277**(1):92-101.
97. Kotov AA, Olenkina OM, Kibanov MV, Olenina LV: **RNA helicase Belle (DDX3) is essential for male germline stem cell maintenance and division in *Drosophila*.** *Biochimica et Biophysica Acta* 2016, **1863**(6 Pt A):1093-1105.
98. Casper AL, Baxter K, Van Doren M: **no child left behind encodes a novel chromatin factor required for germline stem cell maintenance in males but not females.** *Development* 2011, **138**(16):3357-3366.
99. Luebbering N, Charlton-Perkins M, Kumar JP, Lohead PA, Rollmann SM, Cook T, Cleghon V: ***Drosophila* Dyrk2 plays a role in the development of the visual system.** *PLoS One* 2013, **8**(10):e76775.
100. Yoshida S, Yoshida K: **Multiple functions of DYRK2 in cancer and tissue development.** *FEBS Letters* 2019, **593**(21):2953-2965.
101. Wormser O, Levy Y, Bakhrat A, Bonaccorsi S, Graziadio L, Gatti M, AbuMadigham A, McKenney RJ, Okada K, El Riati S *et al*: **Absence of SCAPER causes male infertility in humans and *Drosophila* by modulating microtubule dynamics during meiosis.** *Journal of Medical Genetics* 2020.

102. Kyte J, Doolittle RF: **A simple method for displaying the hydropathic character of a protein.** *Journal of Molecular Biology* 1982, **157**(1):105-132.
103. Sabbia V, Piovani R, Naya H, Rodriguez-Maseda H, Romero H, Musto H: **Trends of amino acid usage in the proteins from the human genome.** *Journal of Biomolecular Structure and Dynamics* 2007, **25**(1):55-59.
104. Pang YL, Abo R, Levine SS, Dedon PC: **Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number.** *Nucleic Acids Research* 2014, **42**(22):e170.
105. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al*: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics* 2012, **28**(12):1647-1649.
106. Langdon WB: **Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks.** *BioData Mining* 2015, **8**(1):1.
107. Rao Y, Wu G, Wang Z, Chai X, Nie Q, Zhang X: **Mutation bias is the driving force of codon usage in the *Gallus gallus* genome.** *DNA Research* 2011, **18**(6):499-512.
108. Guo X, Bao J, Fan L: **Evidence of selectively driven codon usage in rice: implications for GC content evolution of Gramineae genes.** *FEBS Letters* 2007, **581**(5):1015-1021.

Table 1. The organism-wide Δ RSCU values determined using genes with the top 5% expression level (when averaged across all nine tissues) and lowest 5% expression level (* $P < 0.05$ ** $P < 0.001$). The number of putative tRNAs as determined using tRNA-scan and Euk filter [83] are shown. The primary optimal codon per amino acid and its Δ RSCU value are in bold and underlined. An optimal codon that has a relatively high number of tRNAs (≥ 18 , Opt-codon \uparrow tRNAs) and those with no tRNAs, and thus obligately requiring the use of wobble tRNAs (Opt-codon wobble), are shown, as well as the putative wobble anticodon. The primary non-optimal codons that have matching tRNA gene numbers substantially in excess of 0 (≥ 15 , Nonopt-codon \uparrow tRNAs) and those with few/no tRNAs (Nonopt-codon \downarrow tRNAs) are indicated. Codons not having primary optimal or non-optimal status are indicated by "--".

Amino acid	Codon (DNA)	Standard anticodon	Δ RSCU	P	No. tRNAs	Optimal and non-optimal status	Wobble anticodon (optimal)
Ala	<u>GCT</u>	AGC	<u>+0.871</u>	**	35	Opt-codon \uparrow tRNAs	
Ala	GCC	GGC	-0.344	**	0	--	
Ala	GCA	UGC	+0.518	**	18	--	
Ala	GCG	CGC	-1.039	**	22	Nonopt-codon \uparrow tRNAs	
Arg	CGT	ACG	+0.463	**	40	--	
Arg	CGC	GCG	-1.053	**	0	Nonopt-codon \downarrow tRNAs	
Arg	CGA	UCG	+0.185	**	39	--	
Arg	CGG	CCG	-0.548	**	2	--	
Arg	<u>AGA</u>	UCU	<u>+0.881</u>	**	18	Opt-codon \uparrow tRNAs	
Arg	AGG	CCU	+0.047		26	--	
Asn	<u>AAT</u>	AUU	<u>+0.416</u>	**	0	Opt-codon wobble	GUU
Asn	AAC	GUU	-0.244	**	37	Nonopt-codon \uparrow tRNAs	
Asp	<u>GAT</u>	AUC	<u>+0.520</u>	**	0	Opt-codon wobble	GUC
Asp	GAC	GUC	-0.482	**	31	Nonopt-codon \uparrow tRNAs	
Cys	<u>TGT</u>	ACA	<u>+0.368</u>	**	0	Opt-codon wobble	GCA
Cys	TGC	GCA	-0.365	**	38	Nonopt-codon \uparrow tRNAs	
Gln	<u>CAA</u>	UUG	<u>+0.254</u>	**	39	Opt-codon \uparrow tRNAs	
Gln	CAG	CUG	-0.218	**	37	Nonopt-codon \uparrow tRNAs	
Glu	<u>GAA</u>	UUC	<u>+0.496</u>	**	31	Opt-codon \uparrow tRNAs	
Glu	GAG	CUC	-0.480	**	18	Nonopt-codon \uparrow tRNAs	
Gly	<u>GGT</u>	ACC	<u>+0.610</u>	**	0	Opt-codon wobble	GCC
Gly	GGC	GCC	-0.709	**	41	Nonopt-codon \uparrow tRNAs	
Gly	GGA	UCC	+0.483	**	19	--	
Gly	GGG	CCC	-0.383	**	11	--	
His	<u>CAT</u>	AUG	<u>+0.511</u>	**	0	Opt-codon wobble	GUG
His	CAC	GUG	-0.452	**	37	Nonopt-codon \uparrow tRNAs	
Ile	<u>ATT</u>	AAU	<u>+0.603</u>	**	22	Opt-codon \uparrow tRNAs	
Ile	ATC	GAU	-0.452	**	0	Nonopt-codon \downarrow tRNAs	

Ile	ATA	UAU	+0.045		19	--	
Leu	TTA	UAA	+0.537	**	28	Opt-codon [↑] tRNAs	
Leu	TTG	CAA	+0.383	**	16	--	
Leu	CTT	AAG	+0.409	**	39	--	
Leu	CTC	GAG	-0.629	**	0	--	
Leu	CTA	UAG	+0.007		28	--	
Leu	CTG	CAG	-0.692	**	30	Nonopt-codon [↑] tRNAs	
Lys	AAA	UUU	+0.263	**	20	Opt-codon [↑] tRNAs	
Lys	AAG	CUU	-0.160	**	50	Nonopt-codon [↑] tRNAs	
Phe	TTT	AAA	+0.407	**	0	Opt-codon _{wobble}	GAA
Phe	TTC	GAA	-0.265	**	48	Nonopt-codon [↑] tRNAs	
Pro	CCT	AGG	+0.749	**	36	Opt-codon [↑] tRNAs	
Pro	CCC	GGG	-0.359	**	0	--	
Pro	CCA	UGG	+0.483	**	31	--	
Pro	CCG	CGG	-0.843	**	36	Nonopt-codon [↑] tRNAs	
Ser	TCT	AGA	+0.731	**	36	Opt-codon [↑] tRNAs	
Ser	TCC	GGA	-0.208	**	0	--	
Ser	TCA	UGA	+0.493	**	21	--	
Ser	TCG	CGA	-0.723	**	15	Nonopt-codon [↑] tRNAs	
Ser	AGT	ACU	+0.325	**	0	--	
Ser	AGC	GCU	-0.619	**	60	--	
Thr	ACT	AGU	+0.644	**	35	Opt-codon [↑] tRNAs	
Thr	ACC	GGU	-0.223	**	0	--	
Thr	ACA	UGU	+0.493	**	37	--	
Thr	ACG	CGU	-0.873	**	31	Nonopt-codon [↑] tRNAs	
Tyr	TAT	AUA	+0.430	**	0	Opt-codon _{wobble}	GUA
Tyr	TAC	GUA	-0.186	**	43	Nonopt-codon [↑] tRNAs	
Val	GTT	AAC	+0.600	**	26	Opt-codon [↑] tRNAs	
Val	GTC	GAC	-0.394	**	0	--	
Val	GTA	UAC	+0.314	**	30	--	
Val	GTG	CAC	-0.484	**	40	Nonopt-codon [↑] tRNAs	
Amino acids with one codon							
Met	ATG	CAU			43	--	
Trp	TGG	CCA			32	--	
Total tRNAs					1,391		

Table 2. Top predicted GO functional groups for organism-wide highly expressed genes (top 5% expression levels when averaged FPKM across all nine tissues). The top five clusters with the greatest enrichment (abundance) scores are shown. P-values are derived from a modified Fisher's test, where lower values indicate greater enrichment. Data is from DAVID software [88] using those *G. bimaculatus* genes with *D. melanogaster* orthologs (BLASTX $e < 10^{-3}$ [87]).

Enrichment Score: 18.88	P-value
Ribosomal protein	7.30E-31
Cytosolic ribosome	9.00E-11
Enrichment Score: 12.49	
Mitochondrion	3.50E-17
Enrichment Score: 8.98	
Transit peptide: Mitochondrion	4.30E-04
Enrichment Score: 8.39	
Electron transport	1.90E-10
Enrichment Score: 6.49	
Protein folding	2.40E-10

Table 3. Examples of genes that exhibit the top 5% expression levels in the ovaries and top 5% expression levels in the testes (but are not in the top 5% of any other tissue type, Top5_{One tissue}) in *G. bimaculatus* that have elevated use of a non-optimal codon with high tRNAs counts (Nonopt-codon_{↑tRNAs} status; elevated use in this table indicates the RSCU in a gene is ≥ 1.5). The codons include GTG for Val, GGC for Gly, and CTG for Leu (RSCU values ≥ 1.5). Genes are listed that have an identified putative *D. melanogater* (Dmel) ortholog (best match BLASTX $e < 10^{-3}$ [87] and a known gene name at FlyBase [89]).

GB ID	Dmel ID	Gene Name
Ovaries- GTG for Val (RSCU≥ 1.5)		
GBI_17906-RA	FBgn0039889	<i>ADP ribosylation factor-like 4 (Arl4)</i>
GBI_01735-RA	FBgn0261788	<i>Ankyrin 2 (Ank2)</i>
GBI_16610-RA	FBgn0024227	<i>aurora B (aurB)</i>
GBI_20301-RA	FBgn0000182	<i>Bicaudal C (BicC)</i>
GBI_10942-RA	FBgn0024491	<i>Bicoid interacting protein 1 (Bin1)</i>
GBI_05907-RA	FBgn0000337	<i>cinnabar (cn)</i>
GBI_11302-RA	FBgn0030608	<i>Lipid storage droplet-2 (Lsd-2)</i>
GBI_09650-RA	FBgn0031145	<i>Nuclear transport factor-2 (Ntf-2)</i>
GBI_06633-RB	FBgn0031530	<i>Polypeptide GalNAc transferase 2 (Pgant2)</i>
GBI_13292-RA	FBgn0039214	<i>puffyeeye (puf)</i>
GBI_11680-RC	FBgn0004855	<i>RNA polymerase II 15kD subunit (RpIII5)</i>
GBI_13051-RB	FBgn0025697	<i>scavenger receptor acting in neural tissue and majority of rhodopsin is absent (santa-maria)</i>
GBI_03901-RD	FBgn0003312	<i>shadow (sad)</i>
GBI_03557-RA	FBgn0037802	<i>Sirtuin 6 (Sirt6)</i>
GBI_00841-RB	FBgn0003714	<i>technical knockout (tko)</i>
Testes- GTG for Val (RSCU≥ 1.5)		
GBI_00920-RA	FBgn0038984	<i>Adiponectin receptor (AdipoR)</i>
GBI_00615-RA	FBgn0263231	<i>belle (bel)</i>
GBI_03558-RA	FBgn0032820	<i>fructose-1,6-bisphosphatase (fbp)</i>
GBI_04579-RA	FBgn0030268	<i>Kinesin-like protein at 10A (Klp10A)</i>
GBI_09377-RA	FBgn0015754	<i>Lissencephaly-1 (Lis-1)</i>
GBI_12141-RA	FBgn0038167	<i>Lkb1 kinase (Lkb1)</i>
GBI_02406-RA	FBgn0263510	<i>No child left behind (nclb)</i>
GBI_09426-RA	FBgn0261588	<i>pou domain motif 3 (pdm3)</i>
GBI_08602-RA	FBgn0036257	<i>Rho GTPase activating protein at 68F (RhoGAP68F)</i>
GBI_05329-RA	FBgn0032723	<i>short spindle 3 (ssp3)</i>
Ovaries- GGC for Gly (RSCU≥ 1.5)		
GBI_17906-RA	FBgn0039889	<i>ADP ribosylation factor-like 4(Arl4)</i>

GBI_06216-RA	FBgn0031392	<i>Apoptosis inducing factor (AIF)</i>
GBI_20301-RA	FBgn0000182	<i>Bicaudal C (BicC)</i>
GBI_11302-RA	FBgn0030608	<i>Lipid storage droplet-2 (Lsd-2)</i> <i>VAMP-associated protein of 33kDa ortholog A(Vap-33A)</i>
GBI_05398-RA	FBgn0029687	
GBI_09822-RD	FBgn0261458	<i>capulet (capt)</i>
GBI_01828-RA	FBgn0011296	<i>lethal (2) essential for life (l(2)efl)</i>
GBI_10179-RA	FBgn0024841	<i>pterin-4a-carbinolamine dehydratase (pcd)</i>
GBI_13051-RB	FBgn0025697	<i>santa-maria</i>

Testes- GGC for Gly (RSCU>1.5)

GBI_15155-RA	FBgn0016930	<i>Dual-specificity tyrosine phosphorylation-regulated kinase 2 (Dyrk2)</i>
GBI_09377-RA	FBgn0015754	<i>Lissencephaly-1(Lis-1)</i>
GBI_00388-RA	FBgn0010288	<i>Ubiquitin carboxy-terminal hydrolase (Uch)</i>
GBI_09426-RA	FBgn0261588	<i>Pou domain motif 3 (pdm3)</i>
GBI_05329-RA	FBgn0032723	<i>short spindle 3 (ssp3)</i>

Ovaries- CTG for Leu (RSCU>1.5)

GBI_17906-RA	FBgn0039889	<i>ADP ribosylation factor-like 4 (Arl4)</i>
GBI_01735-RA	FBgn0261788	<i>Ankyrin 2 (Ank2)</i>
GBI_06216-RA	FBgn0031392	<i>Apoptosis inducing factor (AIF)</i>
GBI_07513-RA	FBgn0005666	<i>bent (bt)</i>
GBI_20301-RA	FBgn0000182	<i>Bicaudal C (BicC)</i>
GBI_05907-RA	FBgn0000337	<i>cinnabar (cn)</i>
GBI_11302-RA	FBgn0030608	<i>Lipid storage droplet-2 (Lsd-2)</i>
GBI_16524-RA	FBgn0027786	<i>Mitochondrial carrier homolog 1 (Mtch)</i>
GBI_09650-RA	FBgn0031145	<i>Nuclear transport factor-2 (Ntf-2)</i>
GBI_05851-RA	FBgn0003074	<i>Phosphoglucose isomerase (Pgi)</i>
GBI_06633-RB	FBgn0031530	<i>Polypeptide GalNAc transferase 2 (pgant2)</i>
GBI_09582-RA	FBgn0036187	<i>RIO kinase 1 (RIOK1)</i>
GBI_13051-RB	FBgn0025697	<i>santa-maria</i>
GBI_03901-RD	FBgn0003312	<i>shadow (sad)</i>

Testes- CTG for Leu (RSCU>1.5)

GBI_00369-RA	FBgn0003884	<i>Alpha-Tubulin at 84B (alphaTub84B)</i>
GBI_15155-RA	FBgn0016930	<i>Dyrk2</i>
GBI_03558-RA	FBgn0032820	<i>fructose-1,6-bisphosphatase (fbp)</i>
GBI_10438-RA	FBgn0038923	<i>mitochondrial ribosomal protein L35 (mRpL35)</i>
GBI_09426-RA	FBgn0261588	<i>Pou domain motif 3 (pdm3)</i>
GBI_08602-RA	FBgn0036257	<i>Rho GTPase activating protein at 68F (RhoGAP68F)</i>
GBI_05329-RA	FBgn0032723	<i>short spindle 3 (ssp3)</i>

GBI_00450-RA	FBgn0024289	<i>Sorbitol dehydrogenase 1 (Sodh-1)</i>
GBI_14282-RA	FBgn0029763	<i>Ubiquitin specific protease 16/45 (Usp16-45)</i>

Table 4. The average amino acid use of the top 5% expressed genes (Top5_{One-tissue}) in *G. bimaculatus* and 5% lowest expressed genes for the organism-wide analyses (using average expression across all nine tissue types). **indicates P<0.05 using a two tailed t-test. The number of predicted tRNAs in the genome per amino acid are shown. SE- standard error.

Amino acid (AA)	S/C Score	AA Freq. High exp	SE	AA Freq. Low exp	SE	Percent Diff.	P	tRNAs
Gly	1	6.66	0.21	8.71	0.13	-30.70	**	71
Ala	4.76	7.32	0.24	11.54	0.14	-57.72	**	75
Val	12.28	6.73	0.19	6.27	0.08	+6.80	**	96
Ile	16.04	5.70	0.15	2.91	0.04	+49.01	**	41
Leu	16.04	9.07	0.26	8.13	0.10	+10.31	**	141
Ser	17.86	6.75	0.21	7.63	0.11	-12.94	**	132
Thr	21.62	5.16	0.15	5.08	0.07	+1.69		103
Lys	30.14	6.93	0.18	3.53	0.06	+49.08	**	70
Pro	31.8	4.62	0.15	6.95	0.11	-50.40	**	103
Asp	32.72	5.08	0.16	3.83	0.06	+24.64	**	31
Asn	33.72	4.30	0.13	2.68	0.04	+37.70	**	37
Glu	36.48	6.53	0.22	5.09	0.07	+22.08	**	49
Gln	37.48	3.75	0.15	3.49	0.05	+6.92	**	76
Phe	44	4.10	0.10	2.70	0.04	+34.20	**	48
Arg	56.34	5.61	0.15	10.04	0.12	-78.95	**	125
Tyr	57	3.10	0.08	1.87	0.05	+39.53	**	43
Cys	57.16	2.08	0.06	2.51	0.03	-20.75	**	38
His	58.7	2.24	0.07	2.53	0.04	-12.99	**	37
Met	64.68	2.61	0.06	2.32	0.02	+10.93	**	43
Trp	73	1.18	0.03	1.48	0.02	-25.80	**	32

Notes: A negative correlation between S/C score and the frequency of amino acids was observed for highly and lowly expressed genes (Spearman's Ranked R=-0.87 and -0.75, P<10⁻⁷). Further, a positive correlation between the frequency of amino acids and tRNA counts was observed for highly and lowly expressed genes (Spearman's Ranked R=0.65 and 0.74, P<0.05). Percent Diff.= percent difference.

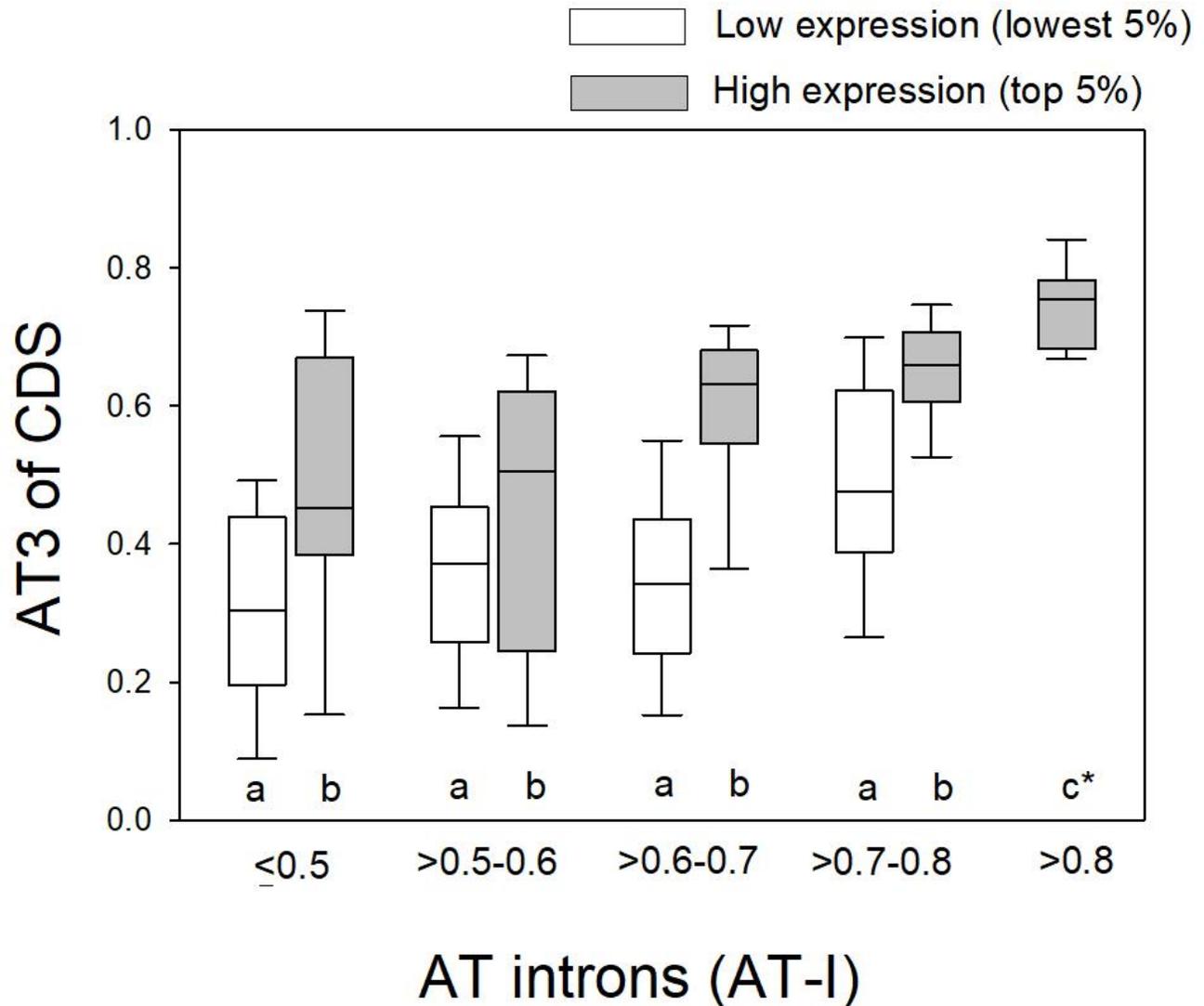


Fig. 1. Box plots of the AT3 of codons of lowly and highly expressed genes within narrow bins of AT-I, and thus presumably having similar background mutational pressures. Genes were binned into categories with similar AT-I content to ascertain differences in AT3 attributable to non-mutational (selection) pressures in highly transcribed genes. Different letters in each pair of bars indicates $P < 0.05$ using MWU-tests. No statistically significant differences in AT-I were observed between highly and lowly expressed genes for any bins (MWU-test $P > 0.30$; with the exception of a minor AT-I difference in medians of 0.019 for category 3 (0.6-0.7)). *AT3 for this bar is statistically significant from all other bars. Only one gene had AT-I > 0.8 for lowly expressed genes and thus the bar for this category was excluded.

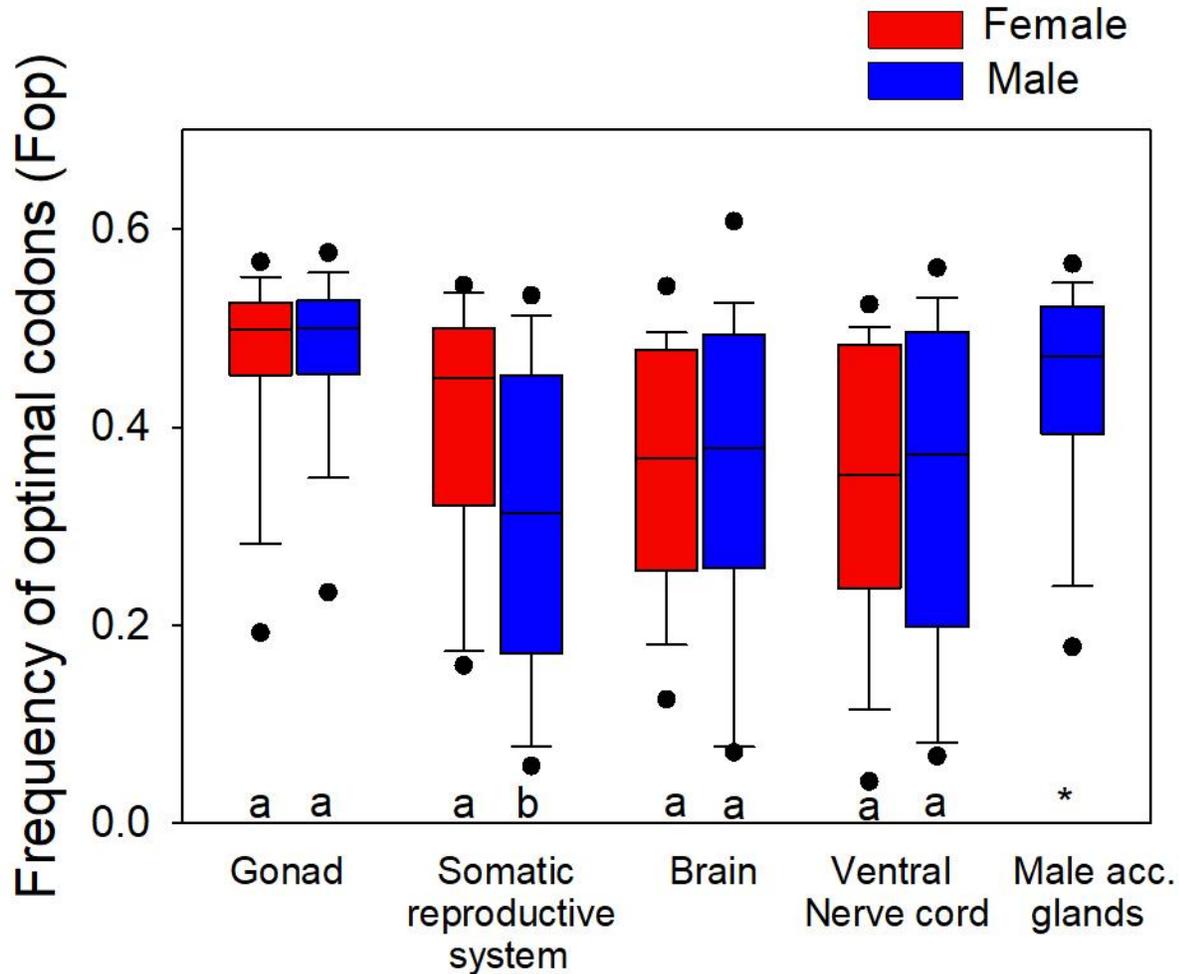
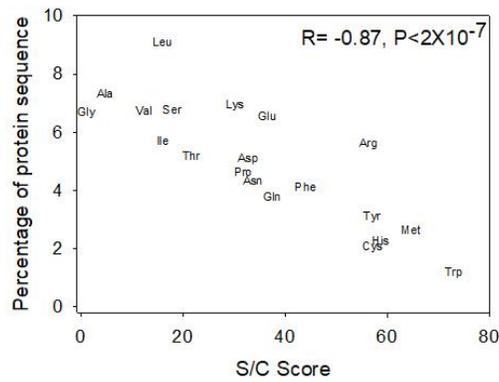
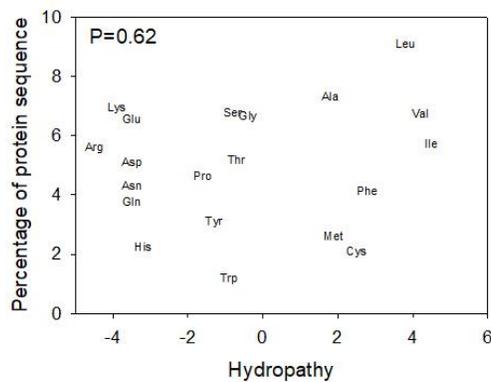


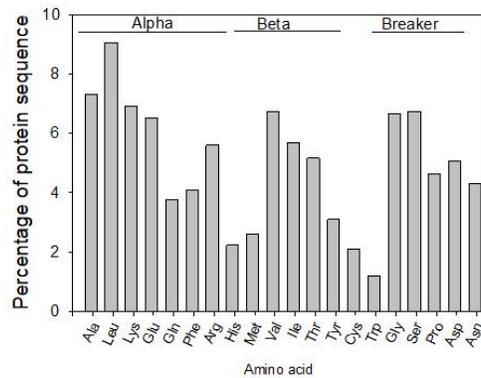
Fig. 2. The frequency of optimal codons (Fop) for genes with expression in the top 5% in one tissue type and not in any other tissues (Top5One-tissue) for *G. bimaculatus*. Different letters within each pair of bars indicates a statistically significant difference (MWU-test $P < 0.05$). Note that the gonad genes had higher Fop values than all other categories (MWU-tests $P < 0.05$). *Indicates a difference of male accessory (acc.) gland genes from all other bars.



A. S/C score, highly expressed genes



B. Hydropathy, highly expressed genes



C. Alpha, beta, breaker; highly expressed genes

Fig. 3. The relationship between amino acid properties and amino acid use (percent per gene, averaged across genes) in the organism-wide highly expressed genes. A) size/complexity (S/C) score; B) hydropathy, and C) folding properties. For A and B Spearman's R and/or P values are shown, and for C no differences were detected between groups (alpha, beta, and breaker, Ranked ANOVA $P > 0.05$).

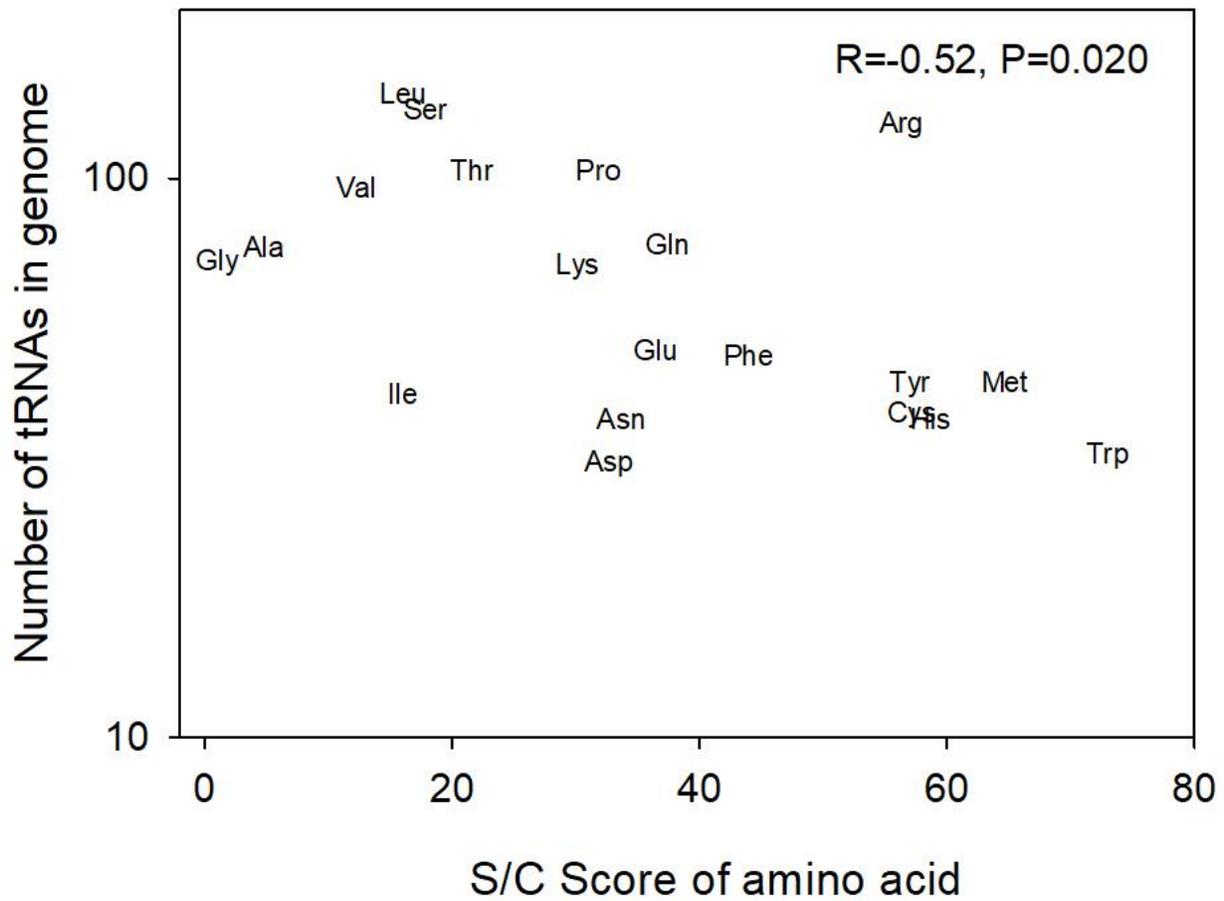


Fig. 4. The predicted gene counts of tRNAs in the *G. bimaculatus* genome and the S/C score of each of 20 amino acids [56]. The Spearman R correlation and P value is shown.