Analysis of Codon Bias in Insect Developmental Genes

Matthew Makowski, David Carlini
Department of Biology
University Honors, Spring 2012

ABSTRACT

Although codon bias, the unequal usage of synonymous codons, has been well documented in the genomes of various organisms, research regarding the presence of codon bias in genes that are maximally expressed during developmental growth stages is less complete. This project examined selected insect species for the presence of codon bias and defined the set of preferred codons in each genome based on patterns of observed codon bias. Also, holometabolous and hemimetabolous insect groups were compared for differences in bias levels. Generally, holometabolic insect species exhibited increased frequency of preferred codon usage in developmental genes versus non-developmental genes, while hemimetabolic species showed the opposite trend, indicating holometabolic development has greater selection pressures for efficient translation during growth. This study then used published lists of gene ontology classifications for *Drosophila melanogaster* (the fruit fly) to identify *D. melanogaster* gene analogs for ontology analysis and compare preferred codon usage within different developmentally relevant biological processes. For example, genes implicated in anatomical structure development demonstrated increased frequency of preferred codon usage in holometabolic but not hemimetabolic species, further supporting the hypothesis that greater selection pressures for efficient protein translation are present during holometabolic development.

INTRODUCTION

One of the most fundamental theorems in biology states that biological information is transcribed from genomic DNA sequences to intermediate mRNAs before translation into protein products. The genetic code contains 61 codons representing 20 separate amino acids indicating 18 codons are degenerate. These 18 degenerate codons are encoded by two to six different codons, and each codon that encodes that same amino acid is termed 'synonymous'. Though mutations between synonymous codons have no effect on the translated amino acid sequence or three-dimensional protein structure, synonymous codons no not appear equally. The unequal frequency of synonymous codons, a phenomenon that occurs in essentially every studied organism including unicellular prokaryotes, eukaryotes, and multicellular eukaryotes, is referred to as codon bias (Hershberg and Petrov 2008; Akashi 2001).

A number of models have been postulated to explain codon usage bias. Perhaps the best supported currently is the translational selection hypothesis, which argues that codon bias is selected for and maintained because codon bias increases the efficiency and accuracy of protein translation through mRNA-tRNA interactions (Hershberg and Petrov 2008). Generally, more frequently appearing synonymous codons, referred to as 'preferred' codons, are those matching the most abundant species of tRNA molecule, indicating a relationship between translation mechanisms and existing codon biases (Ikemura 2001; Kanaya et al. 1999; Kanaya et al. 2001; Yamao et al. 1991; Moriyama and Powell 1997). More importantly, levels of codon bias also correlate strongly with gene expression, as more highly expressed genes exhibit respectively higher levels of codon usage bias (Duret 2002; Gouy and Gautier 1982; Ikemura 1985).

Experimental evidence further supports the translational selection hypothesis. First, preferred codons select their respective aminoacyl-tRNAs more rapidly than rarely used codons do (Curran and Yarus 1989). Second, the introduction of both rare and frequent codons significantly affects the rate of elongation, but incorporation of frequent codons produced a rate approximately six times faster than the rate produced with rare codons (Sorenson et al. 1989). Finally, the replacement of between 1 to 10 preferred codons with unpreferred codons in the alcohol dehydrogenase gene negatively affected both alcohol dehydrogenase expression and *Drosophila* ethanol tolerance, showing codon bias can affect gene function and, conceivably, organism fitness (Carlini 2004; Carlini and Stephen 2003).

Measuring the effect of codon bias in multicellular eukaryotes, relative to much of the early research performed in bacteria and yeast, is more difficult because many eukaryote model organisms undergo distinct developmental phases with potentially variable tRNA pools and selection pressures (Vicario et al. 2008). As such, the body of research concerning codon usage bias in genes maximally expressed during different developmental stages is less complete. First, the observation that the most abundant isoaccepting tRNAs for some amino acids change in different developmental stages led to the discovery that amino acids with unchanged tRNA pools through all developmental stages exhibit increased codon usage bias versus amino acids with variable tRNA pools (Moriyama and Powell 1997). Also, genes expressed at different developmental stages in *Drosophila* show distinctly different levels of codon usage bias, with larval stages showing greatest bias in comparison with embryos, pupae, and adults (Vicario et al. 2008). Vicario et al. assert larvae show maximum codon usage bias because protein content increases rapidly during larval development, approximately a 300-fold increase. As such, high codon usage bias is selected for most strongly during larval development because of the extreme

demand for translational efficiency and accuracy. In sum, these findings support extension of the translation selection hypothesis to codon bias in developmental genes. In fact, the Vicario study in particular indicates that codon usage bias might be enhanced in developmental genes due to their high selection pressures for translation efficiency during rapid larval growth.

There are different types of development that occur, with developmental metamorphosis being perhaps the most obvious. Holometabolous development involves significant larval growth and a dramatic metamorphic molt from larva to pupa to eventual adult. In contrast, hemimetabolic development has gradual nymph stages in which cuticle is repeatedly shed until full adult structures are formed (Gilbert 2010). There is reason to believe codon usage bias patterns would differ observably between holometabolous and hemimetabolous species. First, holometabolic insect species grow linearly at a much more rapid rate than hemimetabolic insect species do (Cole 1980). Presumably, this increased growth rate could correlate with increased selection pressures for translational efficiency during holometabolic development and, subsequently, relatively greater codon usage bias in holometabolic development genes. Second, cuticle production causes less efficient growth in hemimetabolic insects, and hemimetabolic insects in general exhibit decreased nutritional efficiency and respiration when compared with holometabolic insects (Bernays 1986). Obviously holometabolic insects, therefore, have undergone selection for increased growth and efficiency during development, selection that could reasonably have correlated with increased codon usage bias for translational efficiency of relevant metabolic proteins implicated in development. Finally, hemimetabolic muscle protein synthesis rates are actually higher in adult organisms than they are in developing larva (van Marrewijk et al. 1980). However, as previously stated, protein content increases 300-fold during the larval stage in *Drosophila*, a holometabolic insect, a substantially greater protein content

increase than in any other stage indicating a greater rate of protein synthesis (Vicario et al. 2008). According to the translational selection hypothesis, this result indicates codon bias might actually be higher in non-developmental genes in hemimetabolic insects as protein synthesis rates are higher in adult organisms. The opposite trend, then, should hold true for holometabolic insects, if *Drosophila* is considered a model organism.

This research is intended to provide a comprehensive study of the role of codon usage bias in genes maximally expressed during holometabolic and hemimetabolic development. Because of the rapid protein growth occurring during holometabolic larval development and during larval development in general, rapid growth that presumably increases selection pressure for translational efficiency, we expected holometabolic insects to exhibit measurably higher levels of codon usage bias in developmentally relevant genes. In contrast, because growth rates and protein synthesis rates during hemimetabolic development are markedly less than during holometabolic development indicating decreased selection pressure for translational efficiency, we expected hemimetabolic developmental genes would exhibit less codon usage bias relative to holometabolic developmental genes and perhaps less codon bias even than non-developmental hemimetabolic genes.

MATERIALS AND METHODS

Identification of preferred codons in selected model species

Transcriptomes for each organism were identified as all respective mRNA coding sequences catalogued on NCBI's RefSeq databases (Pruit et al. 2009). Using Wright's method

for calculating the effective number of codons (ENC), an ENC value was calculated for each sequence in the transcriptome. Then, the frequency with which each codon appeared in a particular sequence was plotted against the ENC value for that sequence. For visualization, ENC values were binned and codon frequency means were used for those ENC values (Figure 1.A). Preferred codons, then, were defined as the codons that appeared with high frequency at low ENC values, as a low ENC value indicates lower relative degeneracy in the sequence. Using this method applied to the transcriptomes collected from NCBI RefSeq, preferred codons were defined for each amino acid for each of the seven model organisms studied (Figure 1.B). Using this data, the frequency of preferred codon usage (FOP) for each gene in each organism was calculated.

Identification of developmental gene homologs in selected model species

Because comprehensive lists of genes maximally expressed during development are available only in *D. melanogaster* and *T. castaneum*, it was necessary to identify developmental gene homologs in the other species selected for analysis. In this study, The *Tribolium* Sequencing Consortium's published list of *D. melanogaster* developmental genes was used as a query gene set for identifying homologs using NCBI's online TBLASTX software (The *Tribolium* Sequencing Consortium 2008; Altschul et al. 1990). Later, because the online TBLASTX program does not permit excessively large query sets, the NCBI's Standalone TBLASTX software package was utilized to extract *D. melanogaster* homologs for every gene in the other model species' transcriptomes (Altschul et al. 1990). These more comprehensive gene homolog data sets were analyzed with the gene ontology methods described below.

Categorization of gene homologs within gene ontology classifications

Once transcriptome-wide comparisons had been performed to identify *D. melanogaster* gene homologs, it was possible to perform analysis using gene ontology classifications. As with developmental gene lists, complete gene ontologies are available for only a few model organisms, *D. melanogaster* being one of them (The Gene Ontology Consortium 2000). So, each gene in the transcriptomes of the model species studied was matched with its respective gene ontology terms using an online tool produced by the Bioinformatics Group at the Lewis-Sigler Institute at Princeton University (Boyle et al. 2004). Using this method, it was possible to link nearly every gene in each organism with its homologous gene ontology term, or terms, in *D. melanogaster* and produce ontological classifications for nearly every gene.

RESULTS

Developmental genes exhibit greater FOP than non-developmental genes in holometabolous but not hemimetabolous species

When using generalized developmental gene sets from the *Tribolium* Sequencing Consortium to identify developmental gene homologs, differential codon usage bias patterns were observed in holometabolous and hemimetabolous species. Holometabolous species including *Tribolium*, *Drosophila*, *Nasonia*, and *Anopheles* all exhibited increased mean FOP usage in developmental genes when compared with all other non-developmental genes (Table 2).

This result can be visualized in Figure 1.A-D, with the observable shift in FOP usage towards higher FOP values in developmental genes indicating higher codon usage bias levels in developmental genes versus non-developmental genes (Figure 1.A-D). The only exception to this trend among holometabolous species was *Apis*, which exhibited higher mean FOP usage in non-developmental genes (Figure 1.E). Hemimetabolous species, in contrast, exhibited precisely the reverse trend. Among *Pediculus* and *Acyrthosiphon*, increased mean FOP usage was observed in non-developmental genes with developmental genes exhibiting lower bias levels (Table 2, Figure 1.F-G).

No GO Term was consistently over- or under-represented according to odds ratios between the most highly biased and least biased genes

Using transcriptome-wide lists of FOP per gene and gene homolog ontology classifications with *Drosophila*, it was possible to calculate an odds ratio between the 5% most highly biased and the 5% least biased genes within each transcriptome. The intention of this comparison was to discern if any GO Terms had statistically significant over- or under-representation of highly biased genes as compared with genes exhibiting very low bias. Three transcriptomes were analyzed with this method: those of *Drosophila*, *Tribolium*, and *Anopheles*. These three organisms were analyzed first because, as they are the closest phylogenetically of the organisms studied, it was expected they would have the highest chance of showing similar GO term odds ratio patterns, presumably because of closer functional homology with *Drosophila* genes originally categorized into ontologies. However, there was little statistical consistency between over- and under-represented GO terms across even these three species (Figure 3.A-C).

The only GO Term statistically over-represented in all three species was ‘Generation of Precursor Metabolites and Energy’. Neither of the two most developmentally relevant GO terms, ‘Anatomical Structure Development’ and ‘Anatomical Structure Formation Involved In Morphogenesis’ were statistically either over- or under-represented in any of the three species (Figure 3.A-C). Therefore, performing odds ratio comparisons between the 5% most highly biased and 5% least biased genes even between three closely related species offered no conclusive or consistent results.

Genes matching the ‘anatomical structure development’ GO term showed increased mean FOP over all other genes in holometabolous but not hemimetabolous species

By grouping all genes with *Drosophila* homologs identified ontologically as functional in ‘anatomical structure development’, it was possible to compare the mean FOP usage in ‘anatomical structure development’ genes against the mean FOP of every other gene in that organism’s transcriptome. In the holometabolous species *Tribolium, Anopheles,* and *Nasonia,* ‘anatomical structure development’ genes were again shown to have increased mean FOP over non-matching genes (Figure 4.A-C). *Apis* was again shown to be the holometabolous exception, exhibiting decreased mean FOP usage in ‘anatomical structure development’ genes (Figure 4.D). As before, hemimetabolous species exhibited decreased ‘anatomical structure development’ FOP usage compared with all other non-matching genes (Figure 4.E-F).

Comprehensive gene ontology classifications concerning developmentally relevant categories generally follow observed holometabolous/hemimetabolous codon usage bias patterns

By comparing codon usage bias between a number of developmentally relevant gene ontology categories and all other non-matching genes, it was possible to make more holistic, specific observations than could be made by indiscriminately grouping all genes as either 'developmental' or 'non-developmental'. Selected GO terms were analyzed based on their potential relevance to organism growth and development. In general, most GO terms matched the patterns previously established, with holometabolous species showing increased FOP usage in developmentally relevant ontological processes and hemimetabolous species showing the opposite trend (Table 2). GO terms such as 'growth', 'anatomical structure development', and 'cell differentiation' showed high bias in holometabolous species (Table 2). Interestingly, the GO term 'translation' showed practically no increased bias in holometabolous species, indicating that translation selection might not occur on the mechanisms of translation itself (Table 2). Finally, when looking at mean FOP usage for all developmentally relevant GO terms combined, it was surprising to note that *Acyrthosiphon* reversed its earlier trend and actually exhibited increased mean FOP usage across selected GO terms, though this increase was relatively small (Table 2).

DISCUSSION

This research strongly establishes that holometabolous species, within the context of the organisms studied, generally exhibit increased codon usage bias in developmental genes. In the context of generalized 'developmental' gene clusters as defined by the *Tribolium* Sequencing Consortium or more specific developmentally relevant biological processes as classified by the Gene Ontology Consortium, holometabolic species almost always exhibited increased codon

usage bias when compared with non-matching or non-categorized genes (Table 1, Table 2). Similarly, this research strongly establishes that hemimetabolous insects display the opposite trend, with observed codon usage bias increased in non-matching or non-developmental genes (Table 1, Table 2).

Clearly, these findings indicate the presence of increased selection pressures producing and maintaining higher codon usage bias in developmentally relevant genes compared with non-developmentally expressed genes in holometabolous species. It is unclear whether this phenomenon results directly from selection for increased translational efficiency and accuracy during holometabolous development (Hershberg and Petrov 2008). The most viable hypothesis seems to be that holometabolous species experience significantly greater growth and protein content production during development resulting in higher selection pressures on translational efficiency for a large number of genes all related to development. It might be beneficial to reconceptualize codon usage bias, then, within a process-oriented framework. In other words, perhaps codon usage bias should be considered in the context of mechanisms as opposed to individual genes. Gene ontology analysis allows for this possibility. As such, selection pressures could be considered at the level of entire biological processes or functionally related proteins, a scale that could result in more significant fitness changes that might have been invisible on the scale of one or a few genes (Hudson et al. 2011). However, it is clear that more research applying and extending gene ontology analysis would be necessary to fully support this conclusion.

In contrast with holometabolous insects, it was somewhat surprising to observe the consistency with which hemimetabolous developmental or developmentally related genes expressed lower codon bias than other gene groups. Clearly, selection pressures on translational

efficiency are not strong enough to maintain codon bias even at the level of normal, homeostatically expressed genes. Perhaps the explanation for this result stems from van Marrewijk et al.'s observation that muscle protein synthesis in adult *Locusta migratoria* occurs at a more rapid rate than at any developmental stage (van Marrewijk et al. 1980). If muscle synthesis is considered to be a representative process, perhaps hemimetabolic development simply progresses at a relatively slower, more languorous pace before peaking in adulthood. In general, hemimetabolic development does occur in more gradual steps than holometabolic development does (Gilbert 2010). Of course, muscle cell proliferation rates cannot be considered representative of every biological process in the organism. However, it does invoke the intriguing hypothesis that adult homeostasis, in hemimetabolous species, is the stage at which protein content is maximally produced and turned over, thus explaining why hemimetabolic developmental genes show respectively lower bias.

Furthermore, it should be noted that there were two specific exceptions to the holometabolous/hemimetabolous trend described already. First, *A. pisum* actually exhibited a greater mean FOP in selected developmental GO term genes, in contrast with earlier hemimetabolous patterns. Though it is interesting that the trend should reverse after considering a small subset of developmentally relevant processes, the difference in mean FOP in this case was very small. So, it is difficult to make a definitive conclusion from this result, especially after considering that a smaller and more selective group of developmental GO terms could have easily changed the final result. Second, *A. mellifera* consistently exhibited the opposite trend from other holometabolous species, that is, decreased codon usage bias in developmental or developmentally relevant genes. It is difficult to hypothesize why this result occurred, especially because *Nasonia vitripennis*, a close phylogenetic relative of *Apis*, showed exceptionally strong

codon usage bias in developmental genes. One of the most interesting possibilities relates to *Apis*'s exclusive place among studied organisms as a eusocial hymenopteran. It is unclear what effects the genetics of eusocial development and behavior could have on codon usage bias and molecular evolution; however, because *Apis* was the exclusive exception to the holometabolous/hemimetabolous trend observed in this study and the exclusive eusocial organisms, the correlation merits further investigation.

Finally, the results of this study should be understood within the context of its methods. Though strict measures were taken to extract only the most homologous sequences to *D. melanogaster* after TBLASTX comparisons, software that already promotes deep homologies, it must be remembered that extracted sequences were necessarily homologous in primary sequence only. Simply because a gene is identified as developmentally expressed in *D. melanogaster* does not mean that gene's homolog is still implicated in development. It will be important to reconsider and extend these types of process-based gene ontology analyses in the future as more extensive, experimentally verified lists of biological process ontologies are produced.

LITERATURE CITED

Akashi, H. 2001. Gene Expression and Molecular Evolution. *Curr. Opin. Genet. Dev.* 11: 660-666.

Altschul SF, *et al*. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology.* 215: 403–410.

Bernays EA. 1986. Evolutionary contrast in insects: nutritional advantages of holometabolous development. *Physiological Entomology*. 11: 377-82.

Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G. 2004. GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics.* 20(18): 3710-5.

Carlini DB. 2004. Experimental reduction of codon bias in the Drosophila alcohol dehydrogenase gene results in decreased ethanol tolerance of adult flies. *J. Evol. Biol.* 17: 779-85.

Carlini DB, Stephen W. 2003. In vivo introduction of unpreferred synonymous codons into the *Drosophila* Adh gene results in reduced levels of ADH protein. *Genetics.* 163: 239-43.

Cole BJ. 1980. Growth Ratios in Holometabolous and Hemimetabolous Insects. *Annals of the Entomological Society of America.* 73(4): 489-91.

Curran JF, Yarus M. 1989. Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. *J. Mol. Biol.* 209: 65-77.

Duret L. 2002 Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* 12: 640-49.

The Gene Ontology Consortium. 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25(1): 25-9.

Gilbert S. 2010. Developmental Biology. Massachusetts: Sinauer Associates Inc. p. 551.

Gouy M, Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 10: 7055-74.

Hershberg R, Petrov DA. 2008. Selection on Codon Bias. *Annu. Rev. Genet.* 42: 287-99.

Hudson NJ, Gu Q, Nagaraj SH, Ding Y-S, Dalrymple BP, et al. 2011. Eukaryotic Evolutionary Transitions Are Associated with Extreme Codon Bias in Functionally-Related Proteins. *PLoS ONE.* 6(9): e25457.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2:13-34.

Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. 2001. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translational efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J. Mol. Evol.* 53: 290-98.

Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression leve and species-specific diversity of codon usage based on multivariate analysis. *Gene.* 238: 143-55.

van Maarewijk WJ, Schrikker AEM, Beenakkers AMTh. Contents of Nucleic and Amino Acids and Rate of Protein Synthesis in Developing Flight Muscles of *Locusta migratoria*. *Comp. Biochem. Phys.* 65B: 251-257.

Moriyama EN, Powell JR. 1997. Codon usage bias and tRNA abundance in Drosophila. *J. Mol. Evol.* 5: 514-23.

Pruitt KD, Tatusova T, Klimke W, Maglott DR. 2009. NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Res. Database issue:D32-36.

Sorenson MA, Kurland CG, Pedersen S. 1989. Codon usage determines translation rate in *Escheria coli*. *J. Mol. Biol.* 207:365-77.

Tribolium Genome Sequencing Consortium. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature*. 452: 949-55.

Vicario S, Mason CE, White KP, Powell JR. 2008. Developmental stage and Lovel of Codon Usage Bias in Drosophila. *Mol. Biol. Evol.* 25(11): 2269-2277.

Wright F. 1990. The ‘effective’ number of codons used in a gene. *Gene*. 87: 23-29.

Yamao F, Andachi Y, Muto A, Ikemura T, Osawa S. 1991. Levels of tRNAs in bacterial cells as affected by amino acid usage in proteins. *Nucleic Acids Res*. 19:6119-22.

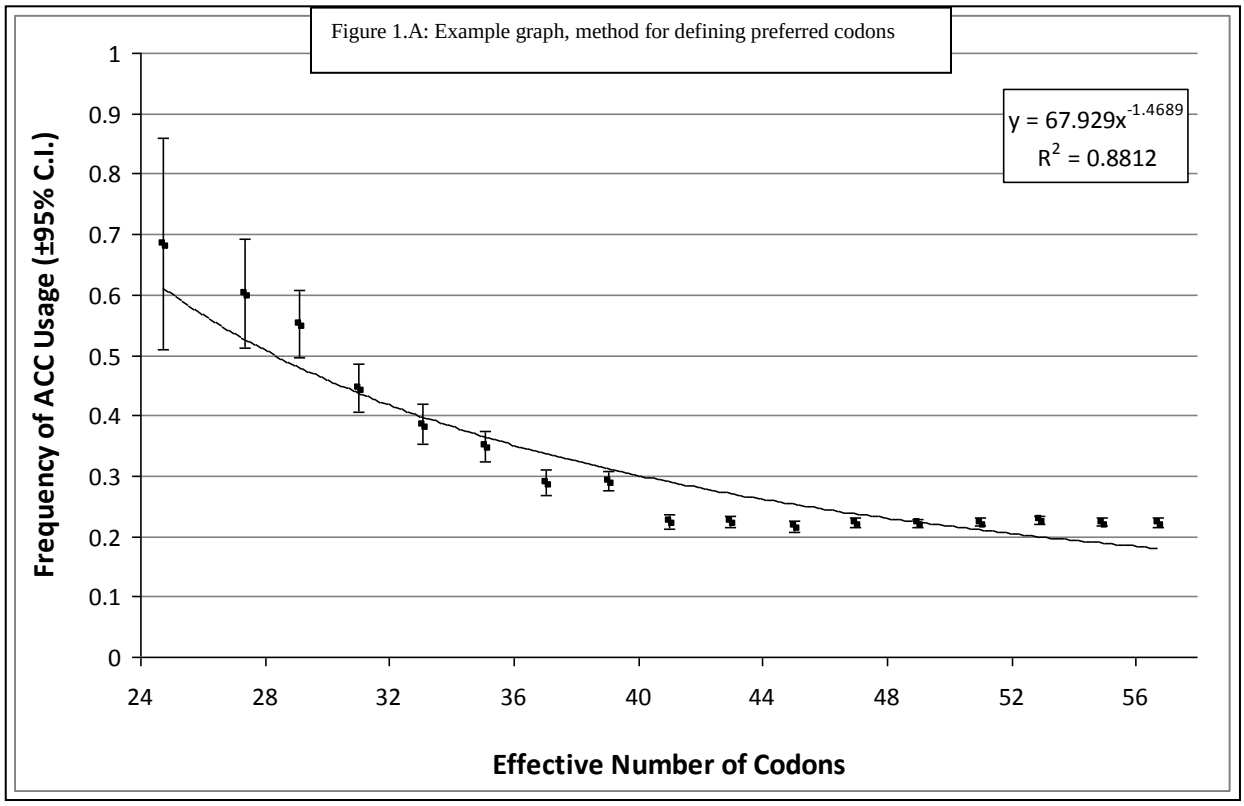


Figure 1.B: All preferred codons as defined for seven holometabolic and hemimetabolic species

		Preferred Codons										Preferred Codons					
		Fruit			Jewel	Flour	Human	Pea				Fruit			Jewel	Flour	Human
Amino		Fly	Mosquito	Honeybee	Wasp	Beetle	Body Louse	Aphid		Amino		Fly	Mosquito	Honeybee	Wasp	Beetle	Body Louse
Acid	Codon	*D. mel*	*A. gam*	*A. mel*	*N. vit*	*T. cas*	*P. hum*	*A. pis*		Acid	Codon	*D. mel*	*A. gam*	*A. mel*	*N. vit*	*T. cas*	*P. hum*
Lys	AAA			**AAA**			**AAA**	**AAA**		Ala	GCA			**GCA**			**GCA**
	AAG	**AAG**	**AAG**		**AAG**	**AAG**					GCG		**GCG**				
Asn	AAC	**AAC**	**AAC**		**AAC**	**AAC**					GCC	**GCC**	**GCC**		**GCC**	**GCC**	
	AAT			**AAT**			**AAT**	**AAT**			GCT			**GCT**			**GCT**
His	CAC	**CAC**	**CAC**		**CAC**	**CAC**				Val	GTA			**GTA**			**GTA**
	CAT			**CAT**			**CAT**	**CAT**			GTG	**GTG**	**GTG**			**GTG**	
Glu	GAA			**GAA**			**GAA**	**GAA**			GTC	**GTC**	**GTC**		**GTC**	**GTC**	
	GAG	**GAG**	**GAG**		**GAG**	**GAG**					GTT			**GTT**			**GTT**
Asp	GAC	**GAC**	**GAC**		**GAC**	**GAC**				Pro	CCA			**CCA**			**CCA**
	GAT			**GAT**			**GAT**	**GAT**			CCG		**CCG**		**CCG**	**CCG**	
Gln	CAA			**CAA**				**CAA**			CCC	**CCC**			**CCC**	**CCC**	
	CAG	**CAG**	**CAG**		**CAG**	**CAG**					CCT			**CCT**			**CCT**
Cys	TGC	**TGC**	**TGC**		**TGC**	**TGC**				Arg	AGA			**AGA**			**AGA**
	TGT			**TGT**			**TGT**	**TGT**			AGG				**AGG**	**AGG**	
Phe	TTC	**TTC**	**TTC**		**TTC**	**TTC**					CGA						
	TTT			**TTT**			**TTT**	**TTT**			CGG		**CGG**		**CGG**		
Tyr	TAC	**TAC**	**TAC**		**TAC**	**TAC**					CGC	**CGC**	**CGC**		**CGC**	**CGC**	
	TAT			**TAT**			**TAT**	**TAT**			CGT	**CGT**		**CGT**			
Ile	ATA			**ATA**						Leu	TTA			**TTA**			**TTA**
	ATC	**ATC**	**ATC**		**ATC**	**ATC**					TTG						
	ATT			**ATT**			**ATT**	**ATT**			CTA						
STOP	TAA		**TAA**	**TAA**			**TAA**	**TAA**			CTG	**CTG**	**CTG**		**CTG**	**CTG**	
	TAG										CTC	**CTC**	**CTC**		**CTC**	**CTC**	
	TGA										CTT			**CTT**			
Thr	ACA			**ACA**			**ACA**	**ACA**		Ser	AGC		**AGC**		**AGC**	**AGC**	
	ACG		**ACG**								AGT			**AGT**			**AGT**
	ACC	**ACC**	**ACC**		**ACC**	**ACC**					TCA			**TCA**			**TCA**
	ACT			**ACT**			**ACT**	**ACT**			TCG	**TCG**	**TCG**				
Gly	GGA			**GGA**			**GGA**	**GGA**			TCC	**TCC**	**TCC**		**TCC**	**TCC**	
	GGG		**GGG**			**GGG**					TCT			**TCT**			**TCT**
	GGC	**GGC**	**GGC**		**GGC**	**GGC**				Met	ATG						
	GGT			**GGT**				**GGT**		Trp	TGG						

Figure 1: (A) Example method for defining preferred codons using *D. melanogaster*. Low ENC indicates a lower overall number of codons are being used. The entire transcriptome is considered. (B) All preferred codons as defined for each model species studied.

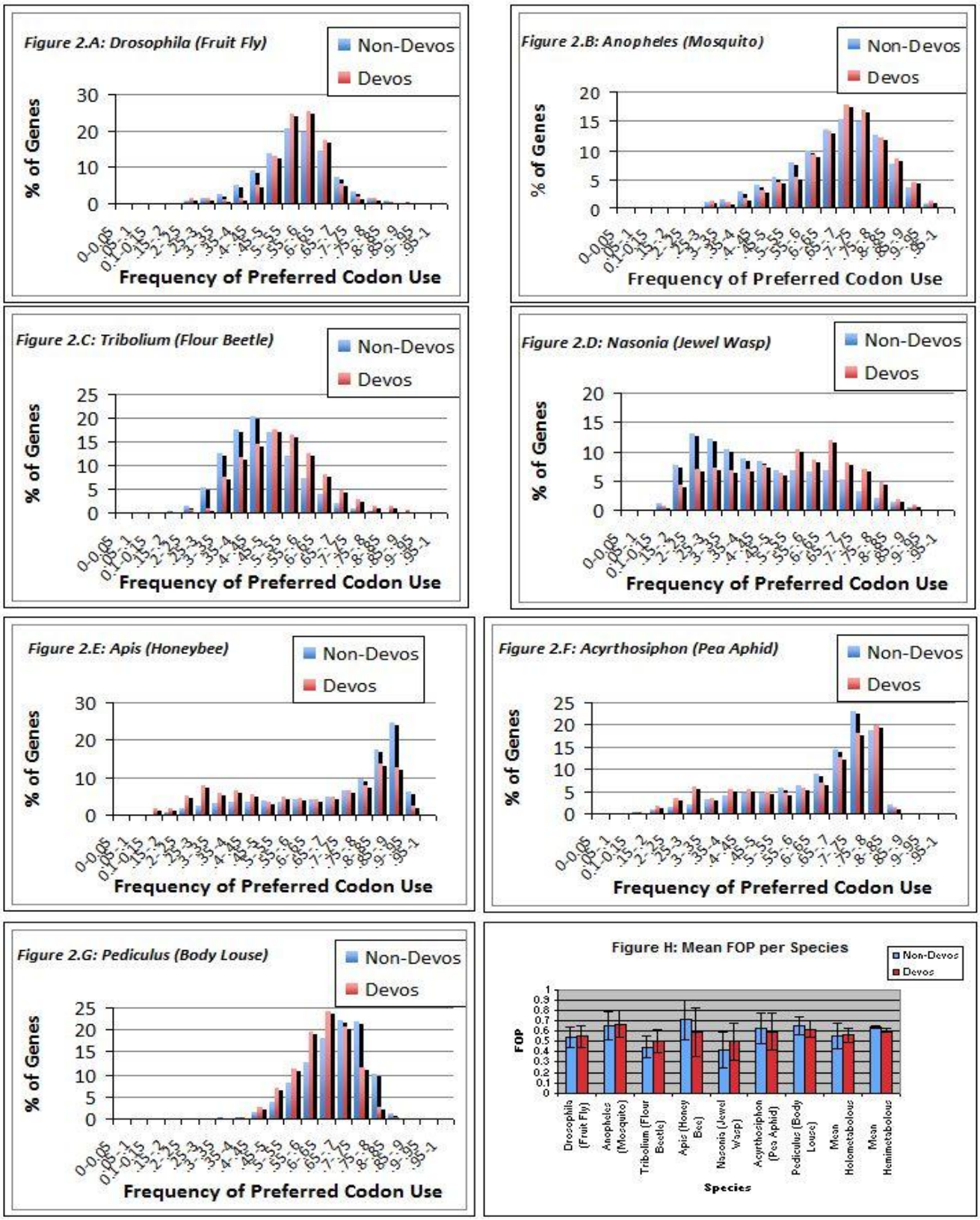
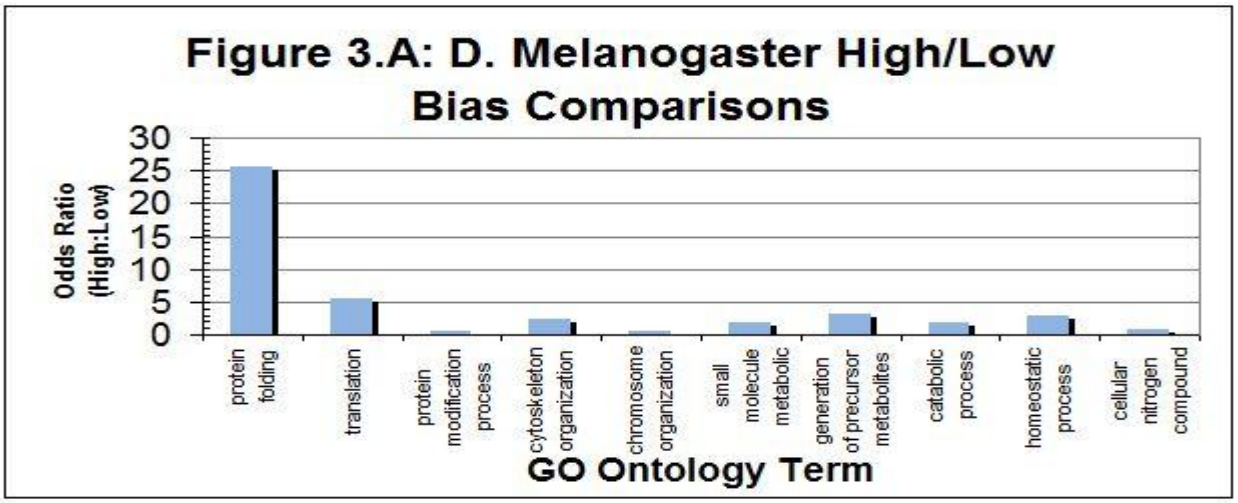
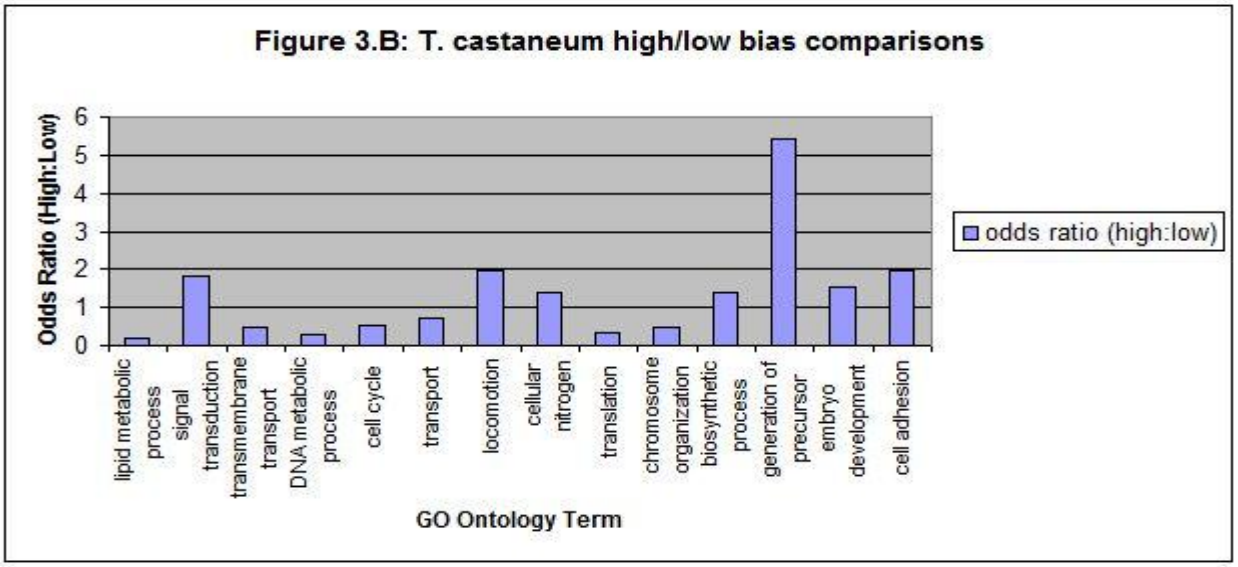


Figure 2: (A-G) Comparison between developmental and non-developmental gene levels of codon usage bias. FOP usage categories are binned for visualization. (H) Mean FOPs are compared per species, with a final comparison for grouped holometabolous and hemimetabolous insects.





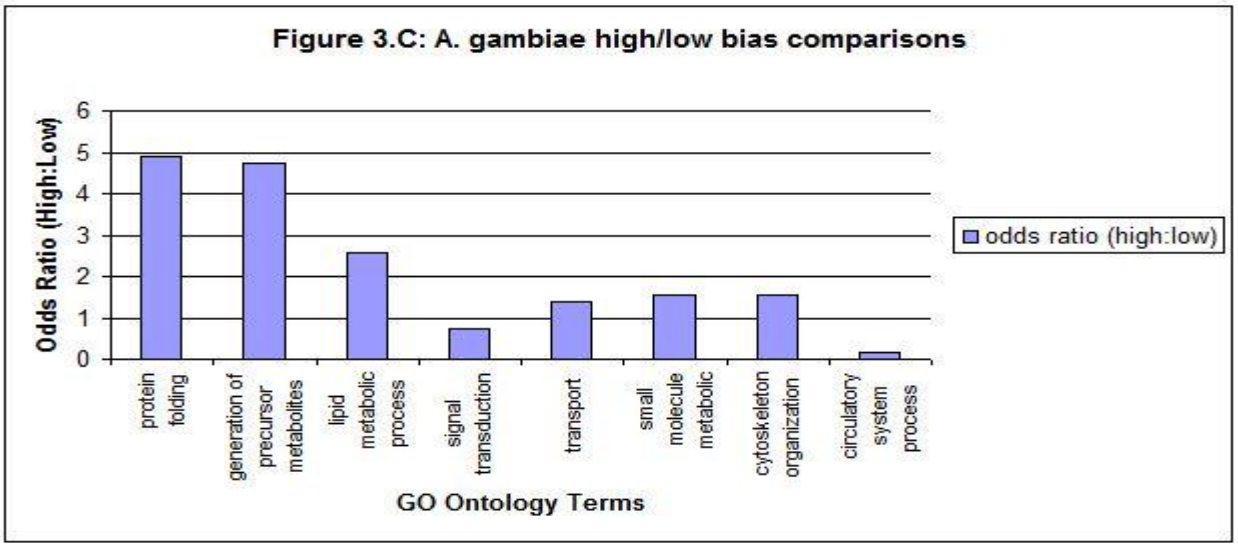


Figure 3: (A-C) Graphs displays the ontology categories with statistically significant odds ratios of highly biased genes to least biased genes, as determined by p-value. Listed GO terms represent either over-representation of highly biased genes (if the ratio is greater than one) or under-representation of highly biased genes (if the ratio is less than one).

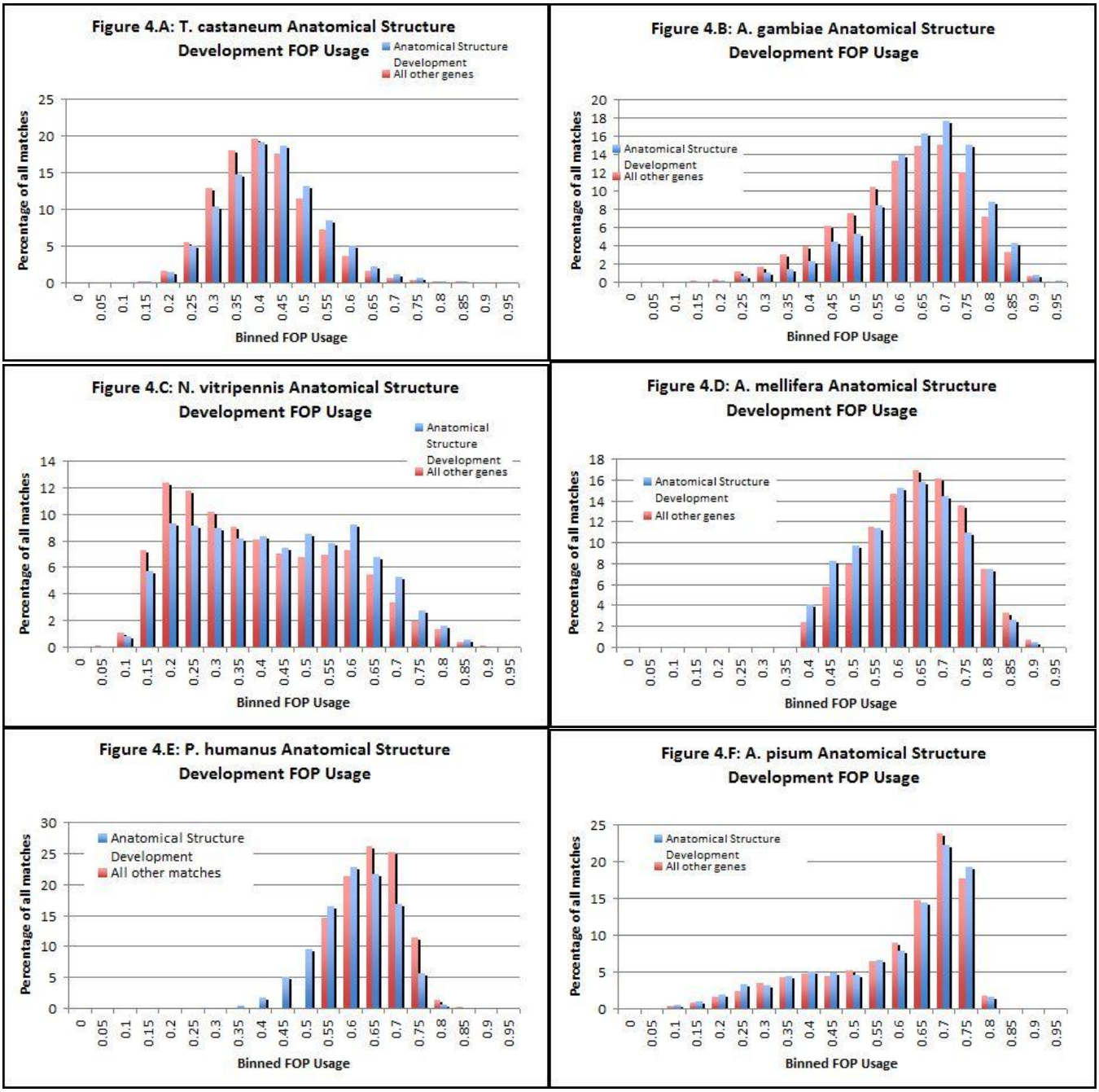


Figure 4: (A-F) Comparison of codon usage bias between the GO term 'Anatomical Structure Development' and all other genes not matching that GO term. Analysis was performed on six species, using *D. melanogaster* gene analogs to identify GO term matches.

Table 1: Comparison of FOP Usage between Developmental and Non-Developmental Genes

Species	**Average FOP Devos**	**Average FOP Non-devos**	**Chi-Square p-value**
Drosophila melanogaster	0.5477	0.5376	3.53E-19
Anopheles gambiae	0.6644	0.6473	0.017
Tribolium castaneum	0.4996	0.4427	2.72E-20
Apis mellifera	0.5891	0.7066	6.26E-21
Nasonia vitripennis	0.4964	0.4129	5.51E-26
Acyrthosiphon pisum	0.5885	0.6247	2.92E-10
Pediculus humanus	0.6146	0.6506	8.24E-39
Mean Holometabolous	0.55944	0.54942	
Mean Hemimetabolous	0.60155	0.63765	

Table 1: Quantifies the FOP usage data visualized in Table 4A-H.

Table 2: Comprehensive FOP Usage Analysis for Selected Developmentally Relevant GO Terms

		T. castaneum			A. gambiae			N. vitripennis			A. mellifera			P. humanus			A. pisum		
		Match Ave FOP	Non-match Ave FOP	t-test	Match Ave FOP	Non-match Ave FOP	t-test	Match Ave FOP	Non-match Ave FOP	t-test	Match Ave FOP	Non-match Ave FOP	t-test	Match Ave FOP	Non-match Ave FOP	t-test	Match Ave FOP	Non-match Ave FOP	t-test
anatomical structure development	GO_0048856	**0.454204956**	0.43621857	6.69E-17	**0.675103**	0.637095	3.08E-35	**0.456177**	0.400602	8.69E-48	0.651472	**0.680998**	2.95E-31	0.632649	**0.653853**	6.86E-34	0.616222	**0.632439**	1.44E-06
cell differentiation	GO_0030154	**0.455959381**	0.43716366	6.76E-17	**0.674214**	0.639959	2.86E-25	**0.456561**	0.404957	1.84E-38	0.649416	**0.679671**	3.71E-30	0.631248	**0.68002**	1.9E-171	0.618384	**0.630179**	0.00078
transport	GO_0006810	**0.452365712**	0.44027959	2.66E-07	**0.678321**	0.641213	1.16E-22	**0.455339**	0.410152	1.16E-26	0.654322	**0.675042**	7.86E-14	0.632113	**0.678947**	1E-136	0.609791	**0.63226**	1.81E-09
signal transduction	GO_0007165	**0.463483186**	0.43831571	3.37E-21	**0.66561**	0.645045	2.96E-06	**0.47308**	0.409414	7.61E-41	0.635013	**0.678424**	3.84E-44	0.62076	**0.678607**	3.2E-156	0.59916	**0.632171**	5.86E-13
cellular component assembly	GO_0022607	**0.451794844**	0.44223919	0.000959	**0.670118**	0.645741	1.13E-05	**0.452317**	0.416468	5.61E-12	0.656991	**0.672041**	6.29E-06	0.631633	**0.676472**	3.47E-76	**0.637909**	0.624531	0.00284
cell morphogenesis	GO_0000902	**0.475235914**	0.4377657	1.88E-35	**0.675614**	0.645578	2.96E-07	**0.483048**	0.411172	9.48E-40	0.62625	**0.678413**	1.33E-51	0.619061	**0.678131**	2.8E-140	0.601845	**0.630586**	6.26E-09
embryo development	GO_0009790	**0.475966328**	0.43754574	7.27E-36	**0.666402**	0.64624	0.001433	**0.485775**	0.411008	2.29E-42	0.627133	**0.677932**	1.56E-48	0.617198	**0.677799**	3E-137	0.595982	**0.631543**	2.8E-12
anatomical structure formation involved in morphogenesis	GO_0048646	**0.469232656**	0.43941459	1.04E-20	**0.661107**	0.646681	0.027674	**0.473963**	**0.413471**	2.41E-28	0.631368	**0.676278**	1.55E-35	0.625094	**0.677377**	1.4E-103	0.619901	**0.627415**	0.116655
translation	GO_0006412	0.443609797	**0.443628673**	0.996798	**0.696543**	0.645452	1.24E-12	0.387443	**0.423606**	3.47E-06	**0.714304**	0.66733	6.66E-25	0.665024	**0.674114**	0.013868	**0.672797**	0.623579	3.9E-22
growth	GO_0040007	**0.467628026**	0.44085162	2.36E-14	**0.677934**	0.646127	1.57E-05	**0.478858**	0.415643	5.95E-24	0.632423	**0.674408**	3.61E-25	0.620779	**0.676665**	2.45E-88	0.609508	**0.628139**	0.001562
cell death	GO_0008219	**0.475589853**	0.44001655	2.12E-22	**0.673729**	0.646336	0.000337	**0.47264**	0.416452	7.17E-18	0.63509	**0.673585**	9.24E-20	0.62526	**0.676064**	7.64E-65	0.622232	**0.626855**	0.376224
cell proliferation	GO_0008283	**0.464844964**	0.44126032	1.3E-11	**0.662202**	0.646909	0.048765	**0.445436**	0.419376	3.98E-05	0.650436	**0.672003**	1.69E-07	0.631356	**0.676138**	3.39E-58	**0.626521**	0.626331	0.972398
cell division	GO_0051301	**0.463628657**	0.44157291	1.96E-09	**0.677572**	0.646631	0.00056	**0.452801**	0.418901	4.01E-07	0.648977	**0.671946**	6.86E-08	0.629445	**0.675917**	9.44E-55	0.623976	**0.626595**	0.638765
mitosis	GO_0007067	**0.45439427**	0.44288780	0.004572	**0.680316**	0.646677	0.0003	0.416117	**0.422216**	0.41561	**0.670782**	0.669707	0.824514	0.642519	**0.67505**	3.97E-23	**0.642586**	0.625058	0.002887
developmental maturation	GO_0021700	**0.45595284**	0.44219895	5.28E-05	**0.651598**	0.6472	0.653742	**0.462265**	0.417301	3.39E-13	0.639786	**0.673253**	6.14E-17	0.627432	**0.676313**	1.18E-66	0.61921	**0.627151**	0.133718
protein folding	GO_0006457	**0.449398158**	0.44354452	0.506395	**0.713495**	0.6465	1.34E-06	0.415335	**0.421925**	0.696405	**0.722427**	0.669014	1.15E-10	**0.674605**	0.674196	0.955338	**0.672658**	0.625722	0.000197
tRNA metabolic process	GO_0006399	0.400779952	**0.443979505**	1.45E-06	**0.660927**	0.647194	0.353829	0.330261	**0.42264**	3.01E-07	**0.771607**	0.668883	1.21E-14	**0.690093**	0.67421	0.03543	**0.725526**	0.625645	7.33E-19
Mean Developmental GO Terms FOP Usage		**0.457298206**	0.440522568		**0.674165**	0.645093		**0.446907**	0.415018		0.65987	**0.674055**		0.636251	**0.675287**		**0.630248**	0.628012	