THE UNIVERSITY OF CHICAGO

# ON TAXONOMICALLY RESTRICTED GENES AS THEY RELATE TO TRANSLATION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECOLOGY & EVOLUTION

BY
JOE DYLAN SOSA

CHICAGO, ILLINOIS
JUNE 2026

Copyright © 2026 by Joe Dylan Sosa

To my parents

*Can I do this, and cannot get a crown? Tut, were it farther off, I'll pluck it down.*

*Henry VI*, Part III (Act III, Scene II). William Shakespeare (1564—1616)

THIS moment as I sit alone, yearning and thoughtful,
it seems to me there are other men in other
lands, yearning and thoughtful;
It seems to me I can look over and behold them,
in Germany, Italy, France, Spain—Or far, far
away, in China, or in Russia or India—talking
other dialects;
And it seems to me if I could know those men better,
I should become attached to them, as I do to men
in my own lands,
It seems to me they are as wise, beautiful, benevolent,
as any in my own lands;
O I know we should be brethren and lovers,
I know I should be happy with them.

*Calamus, 23*

Walt Whitman (1819—1892)

# TABLE OF CONTENTS

LIST OF FIGURES . . . . . vii

LIST OF TABLES . . . . . ix

ACKNOWLEDGMENTS . . . . . x

ABSTRACT . . . . . xi

1 **INTRODUCTION** . . . . . 1
- 1.1 The Genetic Code . . . . . 1
- 1.2 Non-Adaptive and Adaptive Evolution . . . . . 3
- 1.3 New Gene Evolution and Mechanisms of Origination . . . . . 5
- 1.4 Genomic Conflict . . . . . 5
- 1.5 Transfer RNAs, Translation, and Translation Error . . . . . 6
- 1.6 Translational Efficiency Hypotheses . . . . . 10

2 **EVOLUTION OF TRANSFER RNA REPERTOIRES** . . . . . 12
- 2.1 **Abstract** . . . . . 12
- 2.2 **Introduction** . . . . . 12
- 2.3 **Results & Discussion** . . . . . 14
  - 2.3.1 Taxonomically restricted genes exhibit unbiased codon patterns incongruous with nuclear-encoded tRNAs . . . . . 14
  - 2.3.2 tRNA gene repertoires rapidly evolve in copy number and isotype . . 21
  - 2.3.3 Genomic location of tRNA duplicates is related to gene age . . . . . 28
  - 2.3.4 Few recently duplicated isodecoders are directionally selected in natural populations . . . . . 30
  - 2.3.5 Substitutions related to structure, function, and age suggest evolution of functionality . . . . . 32
- 2.4 **Materials & Methods** . . . . . 37
  - 2.4.1 Biased synonymous codon usage and tRNA adaptation test . . . . . 37
  - 2.4.2 Genomic sequence retrieval and annotation . . . . . 38
  - 2.4.3 Gene age dating . . . . . 38
  - 2.4.4 Chromosomal movement . . . . . 39
  - 2.4.5 Whole-genome variant calling . . . . . 40
  - 2.4.6 Structural analysis . . . . . 40

3 **SMALL RNA FUNCTION IN RELATION TO SEX** . . . . . 41
- 3.1 **Abstract** . . . . . 41
- 3.2 **Introduction** . . . . . 41
- 3.3 **Results & Discussion** . . . . . 43
  - 3.3.1 Lineage-specific tRNA duplicates demonstrate sex-biased expression and fragmentation . . . . . 43

3.3.2 Some tRNAs may be selected for sex-associated functions . . . . . 48
3.4 **Materials & Methods** . . . . . 53
3.4.1 Statistical tests of neutrality . . . . . 53
3.4.2 tRNA extraction, library preparation, sequencing, and data analyses . 54
3.4.3 Tissue specificity analysis . . . . . 56

4 **TRANSLATIONAL FIDELITY OF GONADAL PROTEINS** . . . . . 57
4.1 **Abstract** . . . . . 57
4.2 **Introduction** . . . . . 58
4.3 **Results & Discussion** . . . . . 59
4.3.1 Landscape of Amino Acid Misincorporations at High Throughput . . 59
4.3.2 Gonadal Mistranslation Rates Are Not Modulated by Translational Selection . . . . . 62
4.3.3 Potential Mechanistic Basis of Mistranslation in Gonads . . . . . 74
4.4 **Methods & Materials** . . . . . 78
4.4.1 Collection of proteomic data for *D. melanogaster* gonads. . . . . 78
4.4.2 Computing mistranslation rates from proteomic datasets . . . . . 78
4.4.3 Codon Usage and Translation Fidelity . . . . . 79
4.4.4 Characterization of translation error types in *D. melanogaster* . . . . 79

5 **CONCLUSION** . . . . . 81
5.1 Transfer RNA Evolutionary Dynamics . . . . . 81
5.2 Translational Conflict Hypothesis . . . . . 82

BIBLIOGRAPHY . . . . . 85

BRIEF CURRICULUM VITÆ . . . . . 100

# LIST OF FIGURES


1.1 Two paradigms of evolution. . . . . 4
1.2 The transfer of genetic information. . . . . 7
1.3 Translation-associated error . . . . . 8

2.1 Recently derived *D. melanogaster* genes are poorly adapted to the nuclear encoded tRNA pool. . . . . 16
2.2 Ancient genes' codon usage is biased toward the genomic tRNA pool . . . . . 18
2.3 tRNA adaptation is reduced in younger genes . . . . . 19
2.4 tRNA copy numbers exceed expectated values . . . . . 20
2.5 tRNA gene copy number variation across 24 Drosophilid species . . . . . 23
2.6 tRNA gene origination as a rapid, dynamic process . . . . . 25
2.7 tRNA gene age inference . . . . . 26
2.8 *D. melanogaster*-specific isodecoder . . . . . 27
2.9 Nucleotide diversity at tRNA loci . . . . . 31
2.10 tRNA secondary and tertiary structures . . . . . 33
2.11 Distribution of SNPs within structural features of tRNA . . . . . 34
2.12 SNP density across tRNA structural features . . . . . 35
2.13 Predicted tertiary structure of *tRNA-Glu-CTC-5-1* labeled with structural features 37

3.1 The *D. melanogaster* tRNA repertoire exhibits sex-specific expression, fragmentation, and modification. . . . . 45
3.2 Sex-specific post-transcriptional modifications . . . . . 47
3.3 Sense and antisense tRNA transcripts are sex-biased . . . . . 48
3.4 Significant HKA-like test results for tRNA genes . . . . . 50
3.5 SFS Tests Example . . . . . 51
3.6 Evidence of natural selection of sex-biased, young tRNA genes . . . . . 52
3.7 Summary of tRNA evolution evidence . . . . . 53

4.1 Heatmap of ovary codon-specific error detection rates per misincorporated amino acid. . . . . 60
4.2 Heatmap of testis codon-specific error detection rates per misincorporated amino acid. . . . . 61
4.3 Highly expressed genes in ovaries are more codon biased . . . . . 62
4.4 Expression does not predict error rate . . . . . 63
4.5 Increased tRNA copy number does not reduce misincorporation rates. . . . . 70
4.6 Biased synonymous codon usage does predict mistranslation error rates . . . . . 71
4.7 Comparison of expression between proteins with detected and non-detected mistranslation events . . . . . 73
4.8 Mistranslation event according to nucleotides mispaired . . . . . 75
4.9 Tissue-specific error rate per amino acid . . . . . 76
4.10 Age of proteins detected in ovary proteomic datasets . . . . . 77
4.11 Age of proteins detected in testis proteomic datasets . . . . . 78

5.1 Theoretical model of metazoan tRNA evolution. . . . . 83

# LIST OF TABLES

1.1 Standard Genetic Code . . . . . 3
1.2 Summary of Non-Canonical Functions of Transfer RNA Genes . . . . . 8

2.1 Drosophilid species accessions . . . . . 21
2.2 tRNA gene age dating . . . . . 28
2.3 tRNA duplication movement . . . . . 30

3.1 List of Drosophilid species dissected . . . . . 44

4.1 Ovarian codon mistranslation rates . . . . . 64
4.2 Testis codon mistranslation rates. . . . . 67

# ACKNOWLEDGMENTS

I thank my middle and high school music teachers for their tough love and for teaching me a thing or two about life: David Edge, Carl Miller, Sandra Laskowski, Paula Tumlinson Smolik, and Genie Lynn Fuhrman. I thank my undergraduate advisor Dr. Charles Hauser for his for encouragement and for guiding me toward computational biology. Several other professors also made a lasting impact on my undergraduate journey through their teaching and mentorship: Dr. Lisa Goering, Dr. Jason Callahan, Dr. Edward Early, Dr. Casey Sherman, Dr. Michael Kart, and Dr. Laura Baker. I thank my master's advisor Dr. Dapeng Zhang for introducing me to evolutionary biology and for making the great recommendation to consider Chicago for doctoral studies. I give a heartfelt thanks to my advisor Dr. Manyuan Long for making the past several years a wonderful experience and for giving me the freedom to fail and succeed. I thank my committee for their guidance and encouragement: Dr. Marcus Kronforst, Dr. Xin He, Dr. John Novembre, and Dr. Urs Schmidt-Ott. Thanks to all Long Lab members past and present for creating a wonderful environment to train and learn in: Shengqian Xia, UnJin Lee, Emily Mortola, Patricia Zito, Yuan Huang , Chuanzhu Fan, Chuan Dong, Gustavo Rangel, Deanna Arsala, Jianhai Chen, Jared Atlas, Laura Faulere, Natalia Tamarina, Sumaira Shabbir, and Eduardo Dupim. I give a sincere thanks to the Department of Ecology & Evolution including all staff members and faculty. I particularly want to thank Dr. Audrey Aronowsky and Bonnie Brown for shielding me from university bureaucracy and for handling any and all issues I encountered. I thank Dr. Misha Ludwig and Dr. Natalia Tamarina for sharing their good natures with me, increasing my cultural level, and for guiding me through experiments. I also would like to thank Dr. Matthias Steinrücken, Dr. Luís Bettencourt, Dr. Stefano Allesina, and Dr. Greg Dwyer for being great role models and colleagues. Thanks to my family for their support over the many years I've been in school and for never doubting the increasingly esoteric paths I pursued. Finally, thanks to my dear friends who were with me every step of the way.

# ABSTRACT

All extant life converts nucleic acid information to amino acid through protein synthesis. At the same time, genomes are constantly undergoing microevolutionary processes resulting in functional genetic diversity through myriad molecular mechanisms. In this dissertation, I consider the causes and consequences of transfer RNA gene proliferation in eukaryotic species with an emphasis on the role of sex therein. Classically, hypotheses on translational selection and co-evolution of tRNA copy number and codon usage have been invoked to explain the maintenance of tRNA gene duplicates. In contrast, I demonstrate translational conflict between unbiased codon usage of recently evolved genes and the evolving tRNA gene complement. Through comparative genomic analysis, I uncover an unexpectedly rapid rate of tRNA gene turnover across 60 million years of eukaryotic evolution. While new tRNA genes emerge continuously, population genetic analyses determine only a minority of tRNA genes are under directional selection. I empirically investigate the significance of individual tRNA gene copies through small RNA sequencing of reproductive tissues. By comparing expression, fragmentation, and nucleoside modification patterns of individual tRNA copies, I demonstrate that selection of sex-biased functions likely underlies the maintenance of recently duplicated tRNA copies. I then study translational error through computational proteomic analysis of gonadal peptides and present evidence of equivalent mistranslation error rates between optimal and non-optimal codons. Together, I present a model of eukaryotic tRNA repertoire evolution in which nearly neutral processes, sexual selection, and directional selection together effect tRNA functional and copy number variation in metazoans. These findings expand our understanding of how origination time and sex differences affect the evolution and function of eukaryotic non coding RNA gene families. Thus, in addition to multiple tRNA copies supporting translational efficiency, new isodecoding and isoaccepting tRNA copies and their fragments can contribute to maximizing the fitness of each sex.

# CHAPTER I

# INTRODUCTION

Evolution at the molecular level can be understood as the variation in DNA sequences over time within populations of living organisms. Even now, microevolutionary processes are directing the fate of molecules, alleles, and species. As a result of these processes, genes and genomes are continuously being born and lost. Geneticists use letters [A,C,G,T] as shorthand to refer to the basic units of DNA: nucleotides. It's the assortment and recombination of these sequences of adenosine, cytosine, guanine, and thymine which encode biological information. A majority of all sequences in genomes are non-coding, meaning they do not encode proteins. However, these non-coding regions can have important regulatory functions, and some are transcribed into functional RNA molecules such as tRNAs, rRNAs, and microRNAs [Pan, 2018]. In fact, tRNA is one of the largest non-coding gene families in all domains of life, and is essential for the translation of mRNA into proteins [Giege, 2008]. Why do eukaryotic genomes contain hundreds of identical and nearly identical tRNA genes? How did they evolve and what are they doing? All known forms of life require tRNA to accomplish the fundamental processes of mRNA translation and protein synthesis [Fujishima and Kanai, 2014]. However, the functional and evolutionary significance of seemingly redundant tRNA duplicates is under-explored [Ayan et al., 2020, Dale and Uhlenbeck, 2005, Ikemura, 1985, Ledoux et al., 2009, Parisien et al., 2013, Wohlgemuth et al., 2011]. In this dissertation, I focus on the tRNA gene family's evolution and the causes and consequences of its expansion in metazoan species with an emphasis on biologcal sex.

## 1.1 The Genetic Code

The central dogma of molecular biology describes the flow of genetic information from DNA to RNA to protein (Fig. 1.1) [Crick, 1970]. Protein synthesis relies on the genetic code,

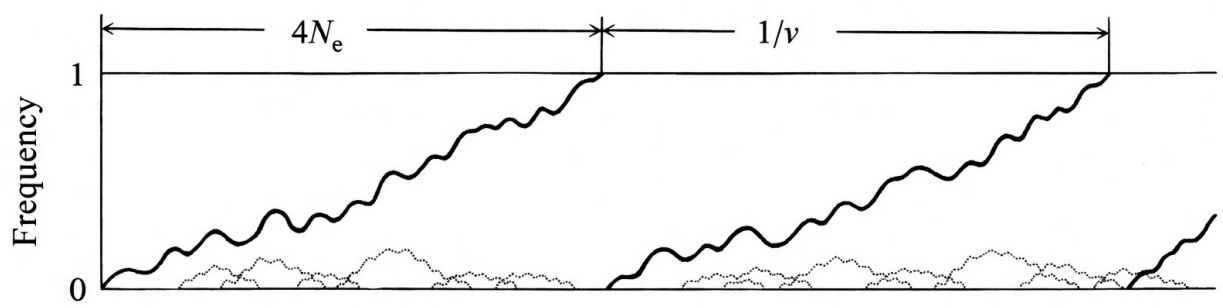
a near-universal triplet codon table that consists of 64 possible codons, derived from the four RNA bases, adenine (A), cytosine (C), guanine (G), and uracil (U), yielding $4^3 = 64$ combinations [Crick, 1968]. In eukaryotes, the 64 mRNA codons correspond to 20 standard amino acids plus selenocysteine and three stop signals. The degeneracy of the genetic code allows multiple codons to specify the same amino acid (Table 1), [Crick, 1958, Nirenberg and Matthaei, 1961, Crick, 1968]. Francis Crick's wobble hypothesis explains how a single tRNA anticodon can recognize multiple synonymous codons through flexible base pairing, particularly at the third (wobble) position, thereby reducing the number of required tRNAs while maintaining decoding accuracy [Crick, 1966, Agris et al., 2018]. No species encodes all possible tRNAs for the 61 sense codons, organisms instead rely on wobble pairing to decode multiple codons with fewer tRNA isoacceptors [Murphy and Ramakrishnan, 2004]. Eukaryotic species in particular universally lack at least 17 cognate tRNAs [Rak et al., 2018], and *D. melanogaster* lacks 18 total (Table. 1).

Table 1.1: Standard Genetic Code. Gray cells indicate codons which lack a corresponding transfer RNA gene in the *D. melanogaster* genome.

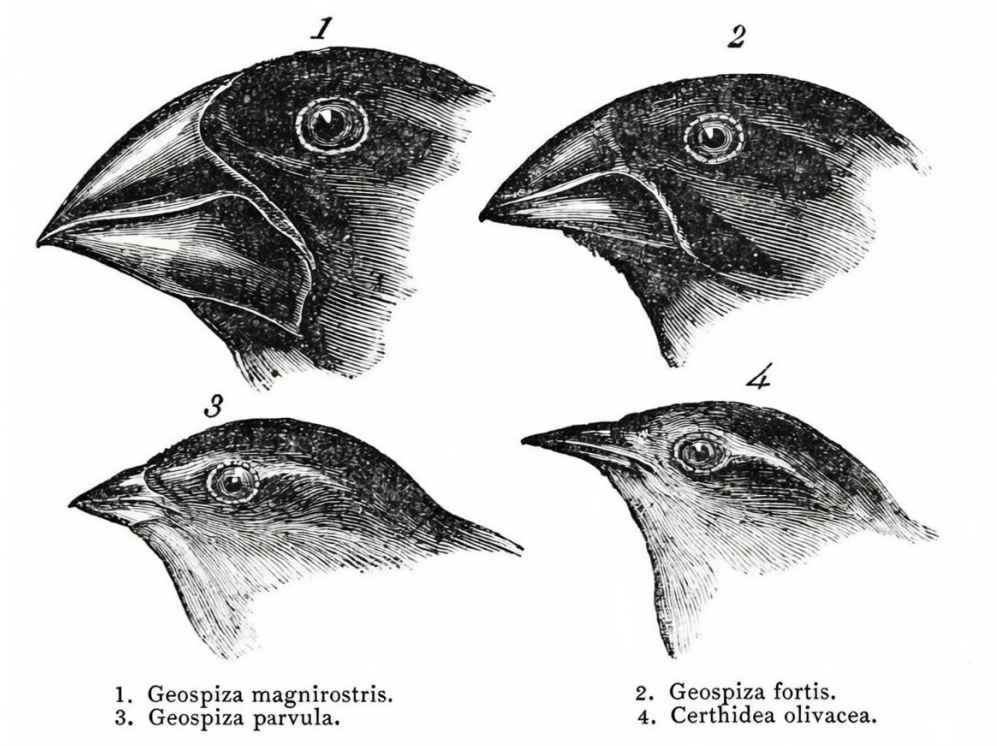
| **1st Base** | **2nd Position** | | | | **3rd Base** |
|---|---|---|---|---|---|
| | **U** | **C** | **A** | **G** | |
| **U** | UUU (Phe) | UCU (Ser) | UAU (Tyr) | UGU (Cys) | **U** |
| | UUC (Phe) | UCC (Ser) | UAC (Tyr) | UGC (Cys) | **C** |
| | UUA (Leu) | UCA (Ser) | UAA (STOP) | UGA (STOP/SeC) | **A** |
| | UUG (Leu) | UCG (Ser) | UAG (STOP) | UGG (Trp) | **G** |
| **C** | CUU (Leu) | CCU (Pro) | CAU (His) | CGU (Arg) | **U** |
| | CUC (Leu) | CCC (Pro) | CAC (His) | CGC (Arg) | **C** |
| | CUA (Leu) | CCA (Pro) | CAA (Gln) | CGA (Arg) | **A** |
| | CUG (Leu) | CCG (Pro) | CAG (Gln) | CGG (Arg) | **G** |
| **A** | AUU (Ile) | ACU (Thr) | AAU (Asn) | UCC (Ser) | **U** |
| | AUC (Ile) | ACC (Thr) | AAC (Asn) | AGC (Ser) | **C** |
| | AUA (Ile) | ACA (Thr) | AAA (Lys) | AGA (Arg) | **A** |
| | AUG (Met) | ACG (Thr) | AAG (Lys) | AGG (Arg) | **G** |
| **G** | GUU (Val) | GCU (Ala) | GAU (Asp) | GGU (Gly) | **U** |
| | GUC(Val) | GCC (Ala) | GAC (Asp) | GGC (Gly) | **C** |
| | GUA (Val) | GCA (Ala) | GAA (Glu) | GGA (Gly) | **A** |
| | GUG (Val) | GCG (Ala) | GAG (Glu) | GGG (Gly) | **G** |

## 1.2 Non-Adaptive and Adaptive Evolution

Molecular evolution includes both adaptive and non-adaptive processes. The neutral theory of molecular evolution posits that most molecular changes are selectively neutral and fixed by genetic drift [Kimura, 1985]. In contrast, adaptive evolution occurs when positive selection fixes beneficial mutations, as seen in the rapid gain of essential functions by new genes in *Drosophila* [Xia et al., 2021, 2025] or in the traits of an organism that are highly suited to their environment (Fig. 1). Codon usage bias reflects a balance: translational selection (adaptive) is thought to favor optimal codons [Akashi, 1994, dos Reis et al., 2004], while non-adaptive processes such as biased gene conversion or mutational pressure also shape synonymous site evolution [Qian and Zhang, 2021].




(a) Bold lines indicate the changes in allele frequency of mutant genes destined for fixation after their birth in a population. $N_e$ is the effective population size while $\nu$ is the mutation rate. Adapted from Kimura [Kimura, 1985].




(b) Illustrations of four species of ground finch from the Galápagos Islands. Adapted from Darwin [Darwin, 1889].

Figure 1.1: Two paradigms of evolution. (a) The nearly neutral theory of molecular evolution posits that the majority of mutations are either deleterious or neutral, with a small fraction being beneficial. (b) The theory of evolution by natural selection explains phenotypic changes in heritable traits in populations as being a result of differential reproductive success of individuals with advantageous genotypes.

## 1.3 New Gene Evolution and Mechanisms of Origination

New gene evolution is a fundamental and continuous process affecting the genomic composition of all organisms [Xia et al., 2025]. Often, newly evolved genes are found only in certain lineages or species. These taxonomically restricted genes arise through diverse molecular mechanisms and contribute significantly to phenotypic innovation and organismal adaptation [Dong et al., 2025]. Retroposition (RNA-mediated duplication) is a prominent mechanism, with many new genes in *Drosophila* originating from retroposed copies that often acquire testis-biased expression [Betran et al., 2002]. Gene duplication followed by neofunctionalization or subfunctionalization remains a major route, as originally proposed by Ohno and later supported by population-genetic and comparative genomic studies [Ohno, 1970, Ulfar Bergthorsson and Roth, 2007]. *De novo* origination from non-coding sequences also plays an important role, particularly in generating lineage-specific proteins with essential developmental functions in *Drosophila* [Zhou et al., 2008, Xia et al., 2021]. Critically, recent work has shown that new genes frequently exhibit rapid evolution of essential roles, subcellular localization patterns, and tissue-specific expression [Chen et al., 2013, Long et al., 2013, Dong et al., 2025, Xia et al., 2025].

## 1.4 Genomic Conflict

Genomic conflict arises when different genetic elements or chromosomal regions have opposing fitness interests, driving rapid evolution and turnover of genes. In *Drosophila*, sexually antagonistic alleles—beneficial in one sex but detrimental in the other—fuel conflict, with many male-biased genes accumulating on autosomes after escaping the X chromosome [Innocenti and Morrow, 2010, Zhang et al., 2010a,b]. New genes can also resolve sexual conflict by acquiring essential gametogenesis functions as reported in [VanKuren and Long, 2018]. Heterochromatin-associated gene families, such as ZAD-ZNF genes, evolve rapidly under

conflict-driven pressures, contributing to species-specific genome defense and chromatin regulation [Kasinathan et al., 2020]. These patterns illustrate how intragenomic conflict shapes gene birth, movement, and functional innovation.

## 1.5 Transfer RNAs, Translation, and Translation Error

Transfer RNA (tRNA) genes encode adapter molecules essential for translation, where each tRNA is charged with a specific amino acid and recognizes cognate codons via anticodon codon base pairing (Fig. 1.3) [Giege, 2008, Caetano-Anolles and Sun, 2014]. Accurate translation depends on tRNA identity elements, modifications, and abundance matching codon demand, ensuring efficient and faithful polypeptide synthesis [Ikemura, 1985, Rak et al., 2018]. However, tRNAs also contribute to translation errors when mismatches occur, particularly at the wobble position, allowing near-cognate decoding and amino acid misincorporation (Table 1.4) [Crick, 1966, Agris et al., 2018]. Stress conditions, oxidative damage, or altered tRNA modifications reduce fidelity, increasing mistranslation rates and proteome variability [Netzer et al., 2009, Novoa et al., 2012, Zaborske et al., 2014]. In metazoans, tissue-specific tRNA expression further modulates error profiles and translational accuracy [Dittmar et al., 2006, Pinkard et al., 2020]. Two recent studies particularly have indicated a preference for non-optimal codons in gonads of *D. melanogaster* new genes [Allen et al., 2022, Whittle and Extavour, 2024], suggesting a relationship between gene age and codon usage. Additionally, tRNAs participate in a wide array of non-canonical functions that extend far beyond translation (Table 1.2) [Su et al., 2020, Giege, 2008, Rak et al., 2018]. Under various cellular stresses, tRNAs are cleaved by endonucleases such as angiogenin to produce tRNA-derived fragments (tRFs), which act as regulatory molecules [Yamasaki et al., 2009, Akiyama and Ivanov, 2023]. These fragments have been implicated in modulating gene expression, suppressing retrotransposon activity, inhibiting apoptosis, protecting cells during osmotic stress, and even contributing to intergenerational inheritance of acquired traits

through sperm-mediated transmission [Luo et al., 2018, Goodarzi et al., 2015, Martinez, 2018, Schorn and et al., 2017, Saikia and Hatzoglou, 2014, Chen et al., 2016, Sharma et al., 2016].

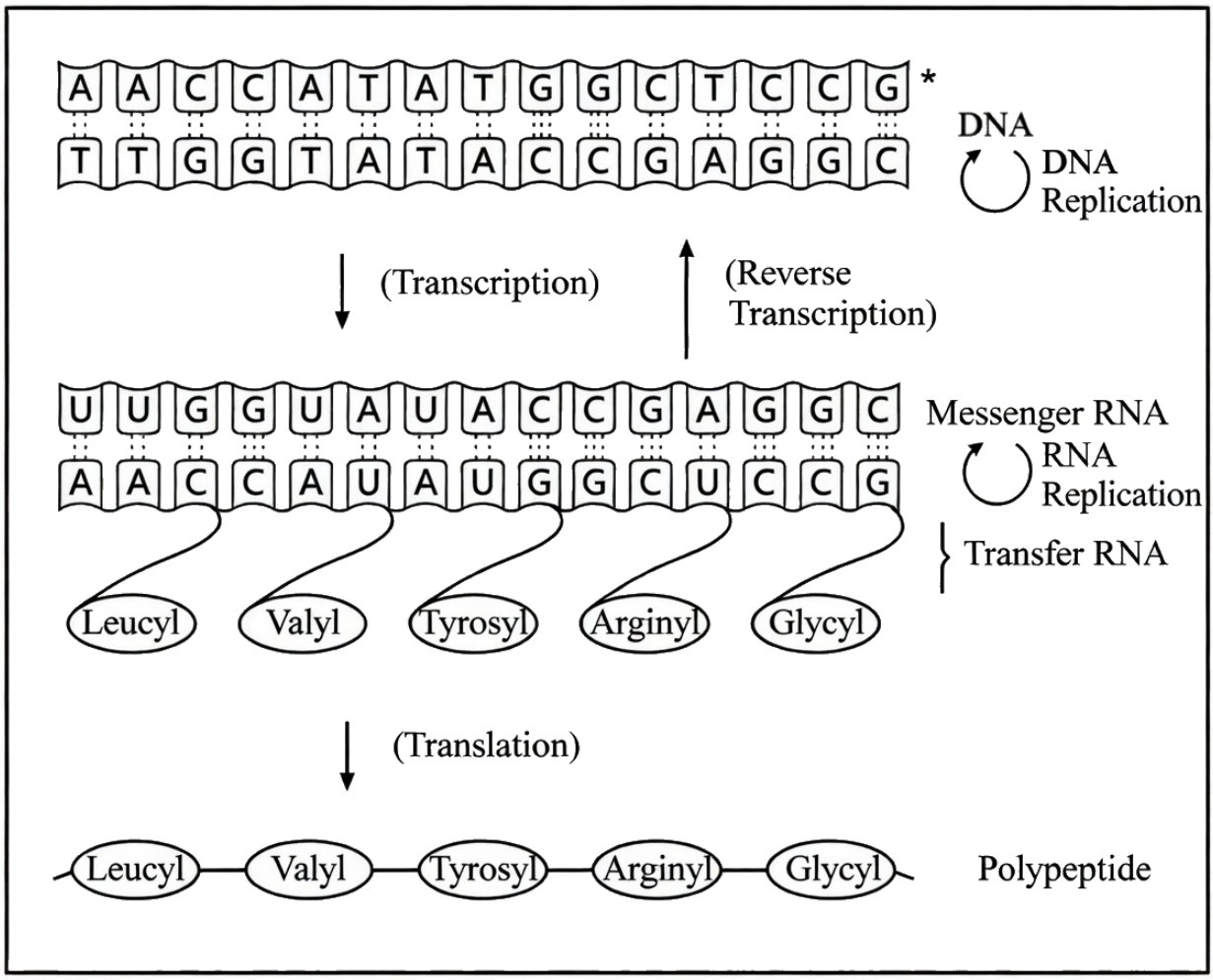



Figure 1.2: Cartoon of DNA transcription, messenger RNA translation, and protein synthesis. Adapted from *Chance and Necessity* by Jacques Monod [Monod, 1971].

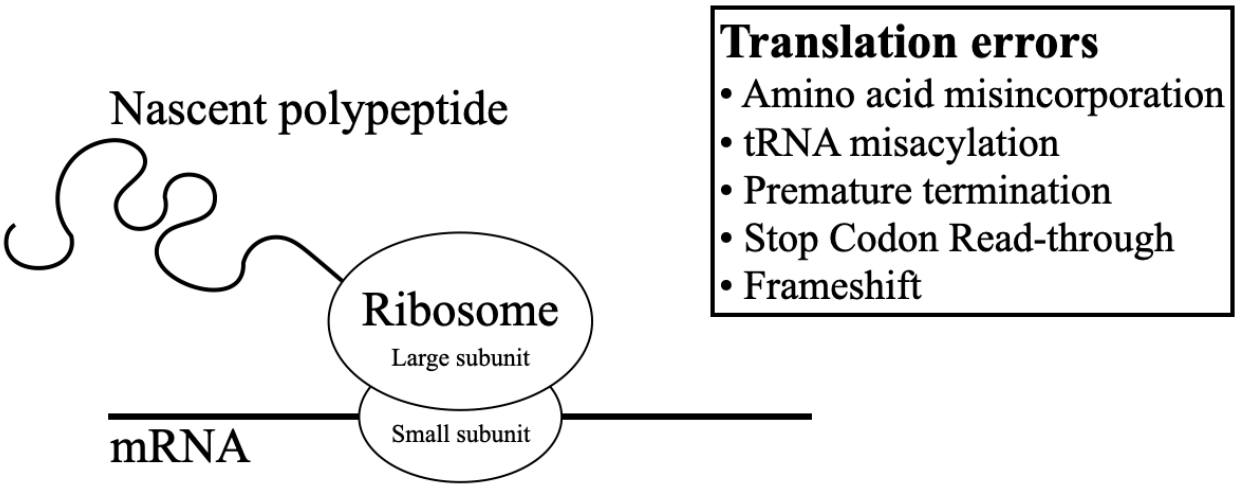



Figure 1.3: An illustration of a ribosome translating an mRNA molecule. Multiple types of errors can occur during mRNA translation, including misincorporation of amino acids, frameshifting, and premature termination.

Table 1.2: Summary of Non-Canonical Functions of Transfer RNA Genes

| **Function** | **Description** | **References** |
|---|---|---|
| Apoptosis regulation | tRNAs bind cytochrome C to inhibit caspase activation and prevent apoptosis. | [Mei et al., 2010, Saikia and Hatzoglou, 2014, Su et al., 2020, Yamasaki et al., 2009] |
| Chromatin organization | tRNA genes function as chromatin insulators and regulate chromatin loops. | [Raab et al., 2012, Rudinger-Thirion et al., 2011, Van Bortle et al., 2017] |
| Extracellular functions | tRNAs and tRFs are secreted in exosomes and biofluids, acting as biomarkers and signaling molecules in intercellular communication and disease. | [Goodarzi et al., 2015, Shurtleff et al., 2017] |

| Function | Description | References |
|---|---|---|
| Immune response modulation | tRNA modifications suppress innate immune receptors, and tRFs regulate viral RNA processing. | [Schlee and Hartmann, 2016, Zhu et al., 2024] |
| Intergenerational epigenetic inheritance | tRFs in sperm and other germ cells mediate the transmission of metabolic and stress-induced phenotypes across generations. | [Chen et al., 2016, Sharma et al., 2016, Zhang et al., 2018] |
| Mistranslation and translational fidelity | tRNAs participate in editing mischarged amino acids and enable adaptive mistranslation under stress for cellular protection. | [An and Musier-Forsyth, 2004, Netzer et al., 2009, Schuntermann and et al., 2024, Wiltrout et al., 2012] |
| Nutrient sensing | Uncharged tRNAs activate kinases like GCN2 or trigger stringent response to sense amino acid availability and regulate gene expression. | [Henkin and Yanofsky, 2002, Masson, 2019] |
| Regulation of RNA/DNA processes | tRNAs influence pre-mRNA splicing, polyadenylation, and DNA replication priming. | [Lu and et al., 2015, Tosar and et al., 2018] |
| Regulation of transposable elements and retrotransposons | tRFs inhibit retrotransposon mobility and tRNAs prime reverse transcription in retroviruses. | [Martinez, 2018, Schorn and et al., 2017] |

| Function | Description | References |
|---|---|---|
| tRNA-derived fragments (tRFs) | Small RNA molecules derived from tRNAs that regulate protein translation, Argonaute-dependent gene silencing, stress responses, and other biological processes. | [Keam and Hutvagner, 2015, Krishna and et al., 2019] |

## 1.6 Translational Efficiency Hypotheses

The translational selection hypothesis posits that natural selection favors synonymous codons that align with the most abundant tRNAs, thereby enhancing translation efficiency, speed, and accuracy. Early evidence showed strong correlations between codon usage bias and tRNA content in both unicellular and multicellular organisms [Ikemura, 1985]. In *Drosophila melanogaster*, codon preferences are thought to be shaped by selection for translational accuracy, with preferred codons reducing translation errors [Akashi, 1994]. Coevolution between codon usage and tRNA abundance provides a mechanistic basis for this selection [Bulmer, 1987], and codon-based evidence for translational selection supports ongoing selection for translational performance in some eukaryotic lineages [Yona et al., 2013, Bénitière et al., 2025].

In Chapter II, I investigate the evolutionary history of tRNA gene repertoires in 24 Drosophilid species through comparative genomic analyses of long-read assemblies. I find evidence of recently evolved tRNA genes and continuous emergence of lineage-specific tRNA genes over evolutionary time. I also explore how selective forces maintain new duplicates, how the diversity of eukaryotic tRNA repertoires is influenced by new coding genes, and how gene age relates to tRNA structure. In Chapter III, I utilize a recently developed tRNA

sequencing method to quantify the abundance, fragmentation, and modification of individual tRNA genes in *D. melanogaster* gonads for the first time. I then test for natural selection at tRNA loci, with particular focus on those displaying differential expression. In Chapter IV, I characterize the landscape of mistranslation events in *D. melanogaster* gonads in order to study translational efficiency in the context of reproductive tissues.

# CHAPTER II

# EVOLUTION OF TRANSFER RNA REPERTOIRES

## 2.1 Abstract

Efforts to study gene family expansion largely consider protein-coding genes to the detriment of our understanding of non-coding RNA gene evolution. Indeed, both genomic and functional studies tend to underappreciate evolutionary forces directing the fate of highly abundant non-coding genes, particularly small RNA genes. Eukaryotic genomes encode hundreds of transfer RNA genes ostensibly to ensure efficient translation. How these seemingly redundant tRNA copies arise and are naturally selected in metazoan genomes and populations remains largely unexplored, owing to challenges in accurate sequencing of tRNA loci and their transcripts. We leveraged long-read genome assemblies of 24 Drosophilid species to infer the origination times of the entire *Drosophila melanogaster* tRNA gene repertoire and found that continuous gain and loss of tRNA duplicates throughout 60 million years of divergence has resulted in rapid taxonomic restriction of isodecoding and isoaccepting copies—even producing one isodecoder specific to *D. melanogaster.* Moreover, we identified patterns of global codon usage, especially in lineage-specific genes, incongruous with translational efficiency hypotheses. From these findings, tRNA gene repertoires may be evolving under other pressures beyond adaptive translation optimizaiton.

## 2.2 Introduction

Chiefly known for their central role in protein synthesis through mRNA translation, transfer RNAs constitute one of the largest non-coding gene families in all domains of life. Most tRNA genes are encoded by the nuclear genome and are organized into isoaccepting groups (isoacceptors) based on their anticodon sequence which specifies the amino acid they are

charged with by cognate aminoacyl-tRNA synthetases. Isoaccepting groups can be further subdivided into isodecoding groups (isodecoders) based on substitutions outside of the anticodon loop, furthering the diversity of the tRNA gene family [Giege, 2008]. Eukaryotic species particularly harbor large and seemingly redundant tRNA gene repertoires-the *Homo sapiens* genome encodes 416 tRNA genes while *Saccharomyces cerevisiae* encodes 275 [Chan and Lowe, 2015]. However, little is understood regarding how identical and nearly identical nuclear tRNA gene copies proliferate and are maintained in eukaryotic genomes and populations.

While the expansion of nuclear encoded tRNA gene pool has been of long-standing general interest, studies have often focused on codon optimality [Hanson and Coller, 2018, Kokate et al., 2021] or the primordial origins of tRNA structure [Caetano-Anolles and Sun, 2014, Fujishima and Kanai, 2014]. Studies of the evolutionary history and principles that shape tRNA gene complements in eukaryotes have been constrained by the highly similar and short length of tRNA genes, a lack of syntenic comparison of tRNA loci between multiple long-read assemblies, and difficulties in high-throughput quantification of tRNAs caused by post-transcriptional base modifications [Hosbach et al., 1980, Tang et al., 2009, Tonzetich et al., 1990, Withers et al., 2006].

The substantial variation in isoacceptor type and number between closely related organisms is generally thought to reflect a co-evolutionary relationship between tRNA availability and mRNA demands [Ikemura, 1985, Qian and Zhang, 2021], but this balance may be less straightforward in specific tissues or disease states in metazoans [Goodarzi et al., 2016]. Given the significant interspecies variability of tRNA gene copy numbers and their critical role in mRNA decoding, we investigated *Drosophila* tRNA gene origination through comparative analyses of long-read genome assemblies. In this chapter, we investigate how selective forces maintain new duplicates, how the diversity of eukaryotic tRNA repertoires is influenced by new coding genes, how gene age relates to tRNA structure and location, and what

role sex has in the differential emergence or function of tRNA duplicates.

## 2.3 Results & Discussion

### *2.3.1 Taxonomically restricted genes exhibit unbiased codon patterns incongruous with nuclear-encoded tRNAs*

Adaptive expansion of the cytosolic tRNA pool through duplication may be an effective strategy to ensure mRNA decoding rather than codon optimization of every gene through point mutations [Ulfar Bergthorsson and Roth, 2007, Yona et al., 2013]. Despite limited phylogenetic conservation, new genes acquired through duplication and other mechanisms (e.g. retrotransposition, de novo) often have biased expression in reproductive and neural tissues [Loppin et al., 2005, Ohno, 1970], integrate into conserved pathways [Ross et al., 2013], and significantly affect organismal fitness [Chen et al., 2013, Kasinathan et al., 2020]. As evolutionarily young genes arise in metazoans some may be unoptimized in their codon usage [Zhang et al., 2005], potentially resulting in discordance between their mRNA and the available pool of tRNAs. To test if tRNA repertoires are generally coevolving to adaptively reflect global codon usage, we first measured the relationship between biased synonymous codon usage, $\psi_g$, and the tRNA adaptation index, $tAI_g$, of 13,785 *D. melanogaster* protein coding genes divided into two groups based on their age (young $<$ 40Ma, ancient $\geq 40Ma$) [Dong et al., 2025]. We observed significant correlations ($p < 0.02$ and $p < 0.001$, Pearson's correlation) between codon usage and adaptation to the genomic tRNA pool for both young ($n = 843$) and ancient genes ($n = 12,942$), with a reduction in both the correlation coefficient $S$ and the coefficient of determination $R^2$ of young genes compared to ancient (Fig. 2.1A,B). We next examined genes stratified by their age and inferred origination mechanism: DNA-based duplication, RNA-based duplication, or lineage-specific orphans [Zhang et al., 2010a]. Ancient genes across mechanism types had significant, positive correlations between

$\psi_g$ and $tAI_g$ ($p \leq 0.004$, Pearson's correlation) as well as elevated $R^2$ compared to young genes of each mechanism type. There was no observed trend between $\psi_g$ and $tAI_g$ amongst young lineage-specific and RNA-based duplicate genes ($p = 0.701$ and $p = 0.0617$, Pearson's correlation), but the codon bias of DNA-based duplicates was significantly correlated ($p = 0.0331, R^2 = 0.01$). In both age groups lineage-specific genes had the weakest $S$ and $R^2$ values ($S \leq 0.28$, $R^2 \leq 0.08$), emphasizing that despite the proliferation of tRNA genes within eukaryotic genomes translational selection is not strongly affecting codon composition in recently derived genes or those putatively originated from intergenic regions (Fig. 2.1C,D).

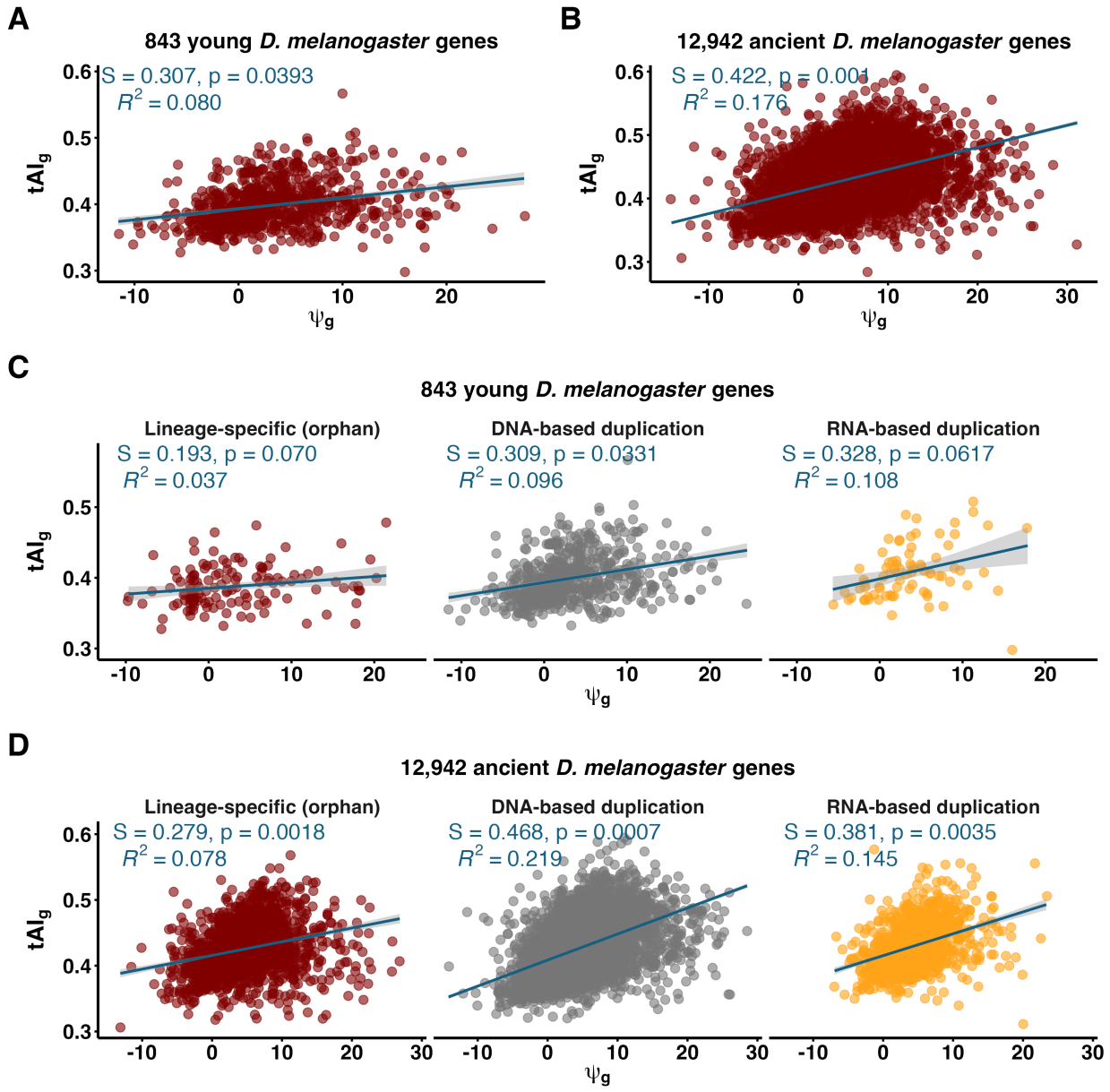



Figure 2.1: **(A)** and **(B)** Test of translation selection using the Pearson correlation ($S$) of biased synonymous codon usage values, $\psi_g$ ,and the tRNA adaptation index, $tAI_g$, of (A) 843 young and (B) 12,942 ancient *D. melanogaster* protein coding genes. Empirical p-values determined by simulation and permutation [Sosa et al., 2025]. **(C)** and **(D)** Same as (A, B) but show lineage-restricted (young) genes and conserved (ancient) genes divided by their inferred origination mechanism.

We then compared the distributions of young and ancient gene $tAI_g$ values for each origination mechanism to interrogate whether unbiased synonymous codon usage was a general

pattern of new gene origination. For each comparison (Fig. 2.2). we observed a significant reduction in $tAI_g$ values of young genes compared to the distribution of ancient genes ($p \leq 1.6 \times 10^{-8}$, rank-sum test). This trend was again observed when splitting the *D. melanogaster* genes into three age groups in order of most distant to most recent divergence time: Drosophila-specific and older, Sophophora subgenus-specific, and melanogaster-subgroup specific. The two younger age groups each had significantly reduced $tAI_g$ values compared to the Drosophila-specific distribution ($p = 2.8 \times 10^{-95}$ and $p = 7.5 \times 10^{-38}$, rank-sum test) and there was a non-significant difference between Sophophora subgenus and melanogaster-subgroup values ($p = 0.053$, rank-sum test) (Fig. 2.3), illustrating a downward tendency of $tAI_g$ as gene age decreases. Although a significant correlation ($R^2 = 0.632$, $p < 0.001$) between amino acid usage frequency and *D. melanogaster* tRNA copy numbers was observed [Bénitière et al., 2025], all isotype copy numbers were distinct from values expected by amino acid usage (Fig. 2.4).

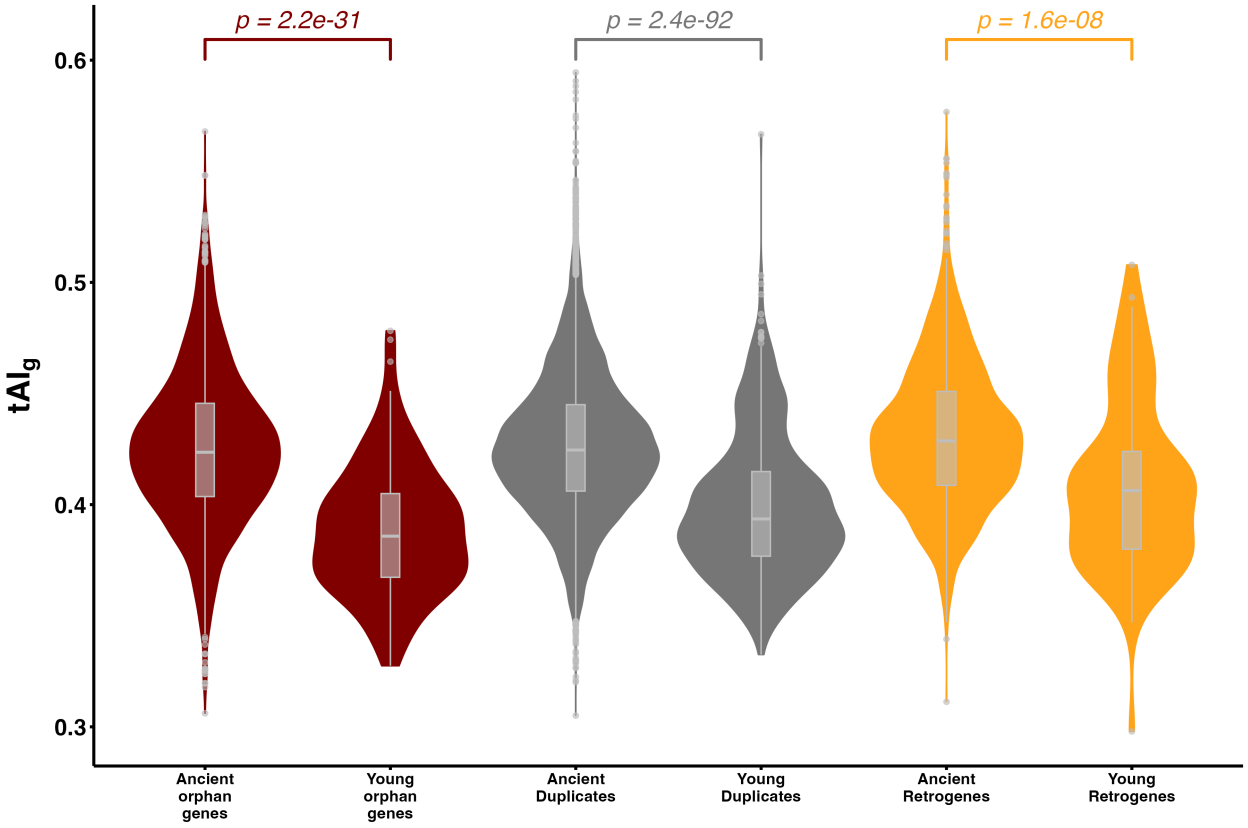



Figure 2.2: Wilcoxon rank-sum test of the distribution of tRNA adaptation index, $tAI_g$, for 13,785 *D. melanogaster* protein coding genes. Distribution of $tAI_g$ grouped by mechanism of gene origination. Overall, ancient genes are better adapted to the genomic tRAN pool.

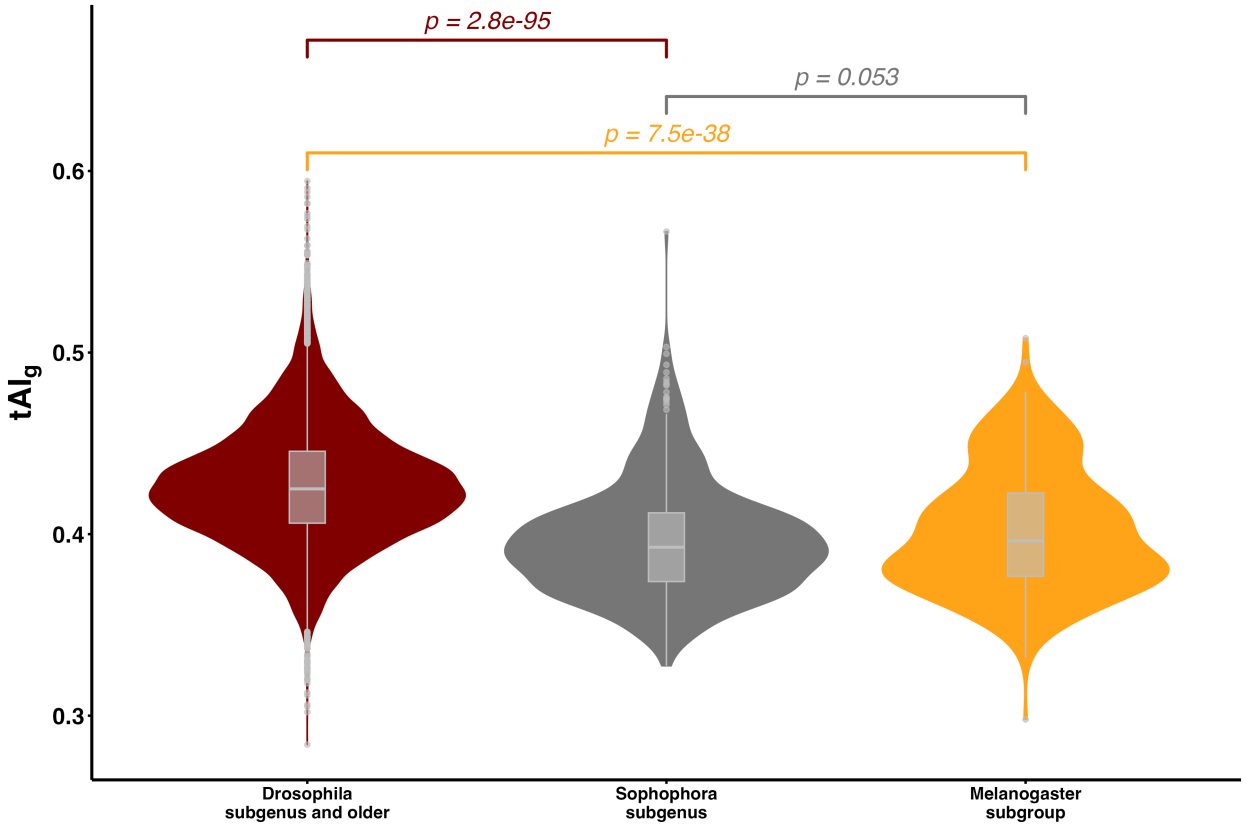



Figure 2.3: Wilcoxon rank-sum test of the distribution of tRNA adaptation index, $tAI_g$, for 13,785 *D. melanogaster* protein coding genes. Distribution of $tAI_g$ for genes divided into three age groups in order of earlier divergence to most recent: *Drosophila* subgenus-specific and older, *Sophohora* subgenus-specific, and *melanogaster* subgroup-specific.

Together, these results reveal that codon usage of recently evolved genes regardless of origination mechanism are maladapted to the nuclear encoded tRNA repertoire, and that tRNA copy numbers exceed or do not meet those predicted by translational efficiency hypotheses. This suggests nuclear encoded tRNA copy numbers are not generally adaptively coevolving with global codon usage and may represent a type of genomic conflict. Rare codons, those non-optimal for translation efficiency, can be necessary for functions of recently derived proteins [Allen et al., 2022] and post-transcriptional remodeling of tRNAs could be a preferred strategy to meet codon usage patterns rather than adaptive duplication [Cope and Shah, 2024, Giguère et al., 2024]. Thus, the tRNA gene family in *D. melanogaster* may be expanding under pressures besides selection for efficient decoding of mRNA.

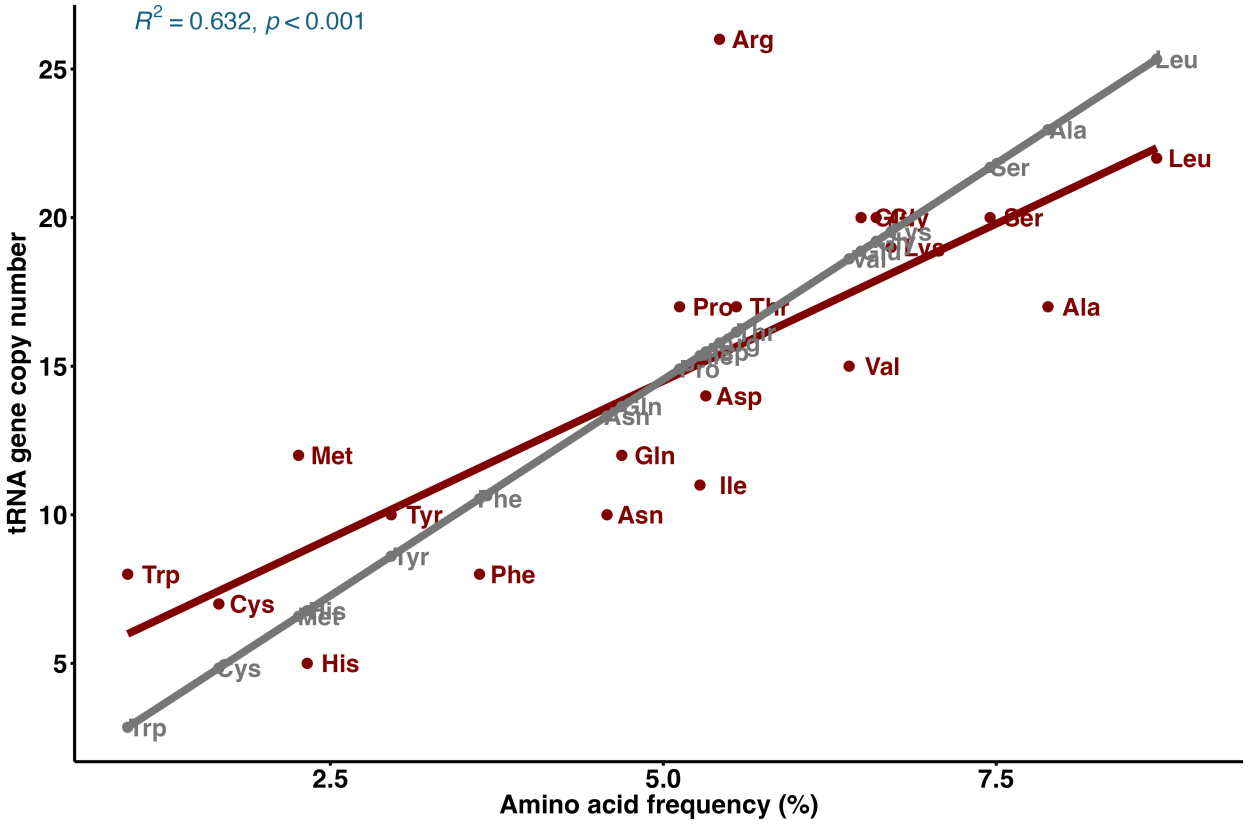



Figure 2.4: Pearson correlations of observed tRNA copy numbers (red) and expected tRNA copy numbers (gray) based on usage frequencies of their corresponding amino acids [Bénitière et al., 2025].

### *2.3.2 tRNA gene repertoires rapidly evolve in copy number and isotype*

Although duplication is known to be a major mechanism by which tRNA genes arise, little is known regarding tRNA gene turnover in closely related species, their duplication patterns, or how tRNA repertoires are maintained in metazoan genomes and populations [Rak et al., 2018]. We first annotated nuclear-encoded tRNA gene sets of 24 Drosophilid species long-read assemblies (Table 2.1) with tRNAscan-SE v.2.0.9 [Chan et al., 2021] and observed highly variable tRNA copy numbers between the 24 closely related species considered in this work (Fig. 2.5A,B). Total Drosophilid tRNA gene content ranged between 241 to 428 genes and unique isodecoding sequences ranged between 88 to 165 per species (Fig. 2.5C). In our focal species, *D. melanogaster*, we annotated 291 nuclear-encoded tRNA genes of which 89 are unique isodecoding sequences. Following these copy number observations, we hypothesized that tRNA gene turnover may occur more frequently in metazoans than expected despite their importance to the central dogma. To test this, we developed a gene age dating methodology to infer the temporal distribution of *D. melanogaster* tRNA gene origination.

Table 2.1: Drosophilid species accessions

| **Species** | **Accession** | **Technology** |
|---|---|---|
| *Scaptodrosophila lebanonensis* | GCF_003285725.1 | PacBio, long-read |
| *Drosophila melanogaster* | GCF_000001215.4 | PacBio, long-read |
| *Drosophila serrata* | GCF_002093755.2 | PacBio, long-read |
| *Drosophila virilis* | GCF_003285735.1 | PacBio, long-read |
| *Drosophila novamexicana* | GCF_003285875.2 | PacBio, long-read |
| *Drosophila hydei* | GCF_003285905.1 | PacBio, long-read |
| *Drosophila persimilis* | GCF_003286085.1 | PacBio, long-read |

Table 2.1 *continued from previous page*

| **Species** | **Accession** | **Technology** |
|---|---|---|
| *Drosophila erecta* | GCF_003286155.1 | PacBio, long-read |
| *Drosophila miranda* | GCF_003369915.1 | PacBio, long-read |
| *Drosophila mauritiana* | GCF_004382145.1 | PacBio, long-read |
| *Drosophla sechellia* | GCF_004382195.2 | PacBio, long-read |
| *Drosophila subobscura* | GCF_008121235.1 | PacBio, long-read |
| *Drosophila pseudoobscura* | GCF_009870125.1 | PacBio, long-read |
| *Drosophila suzukii* | GCF_013340165.1 | PacBio, long-read |
| *Drosophila teissieri* | GCF_016746235.2 | PacBio, long-read |
| *Drosophila santomea* | GCF_016746245.2 | PacBio, long-read |
| *Drosophila yakuba* | GCF_016746365.2 | PacBio, long-read |
| *Drosophila simulans* | GCF_016746395.2 | PacBio, long-read |
| *Drosophila ananassae* | GCF_017639315.1 | PacBio, long-read |
| *Drosophila takahashii* | GCF_018152695.1 | PacBio, long-read |
| *Drosophila grimshawi* | GCF_018153295.1 | PacBio, long-read |
| *Drosophila mojavensis* | GCF_018153725.1 | PacBio, long-read |
| *Drosophila willistoni* | GCF_018902025.1 | PacBio, long-read |
| *Drosophila biarmipes* | GCF_025231255.1 | PacBio, long-read |

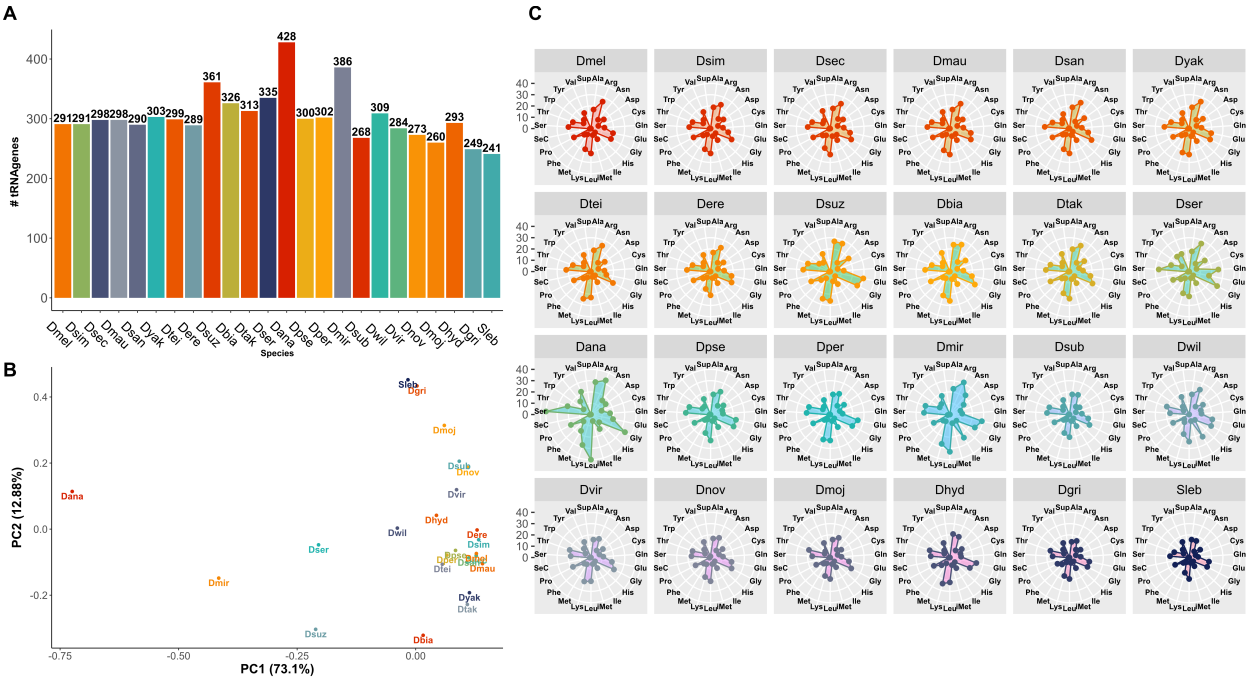



Figure 2.5: tRNA gene copy number variation across 24 Drosophilid species. (A) Total tRNA copy numbers per species considered in this work. Ordered from left to right by increasing time of divergence (rightmost, Sleb, is outgroup). (B) Principal component analysis of isotype and copy number variation in 24 Drosophilids. (C) Radar plots illustrating isotype copy numbers in each species, ordered left to right by row in order of increasing time of divergence. Each white ring indicates an increase of 10 copies, marked on the y-axis 0 through 40.

As tRNA genes are short, largely identical, and often tandemly duplicated, standard phylogenetic or whole genome alignment strategies to determine gene age may be confounded by ambiguous alignments and imprecise homology detection when attempting to investigate their evolution [Holmquist et al., 1973]. Our age dating strategy reduced the homology search space and probability of incorrectly aligning tRNA genes by determining syntenic regions between conserved, neighboring "anchor" protein coding genes. We assigned each *D. melanogaster* tRNA gene to a branch on the Drosophilid divergence time phylogeny [Claudia A. M. Russo and Nei, 1995, Kumar et al., 2017, Obbard et al., 2012] numbered between -1 and 8 (B-1 through B8, larger values indicate more recent divergence) corresponding to the inferred time of their duplication (Fig. 2.6, Table 2.2). We observed the origination of *D.*

*melanogaster* tRNA genes to be a continuous process occurring at a rate of 1.93 gene births per Ma (77 genes / 40 Ma). We found that 26.46% of the tRNA gene repertoire is restricted to the *Sophophora* subgenus (B1 – B8) and 1.03% is further restricted to the *melanogaster* subgroup (B6 - B8). We identified one *D. melanogaster*-specific tRNA isodecoder, *tRNA-Glu-CTC-5-1*, which contains five fixed substitutions compared to outgroup species (Fig. 2.6C), highlighting rapid substitution as previously reported at tRNA loci [Thornlow et al., 2018].

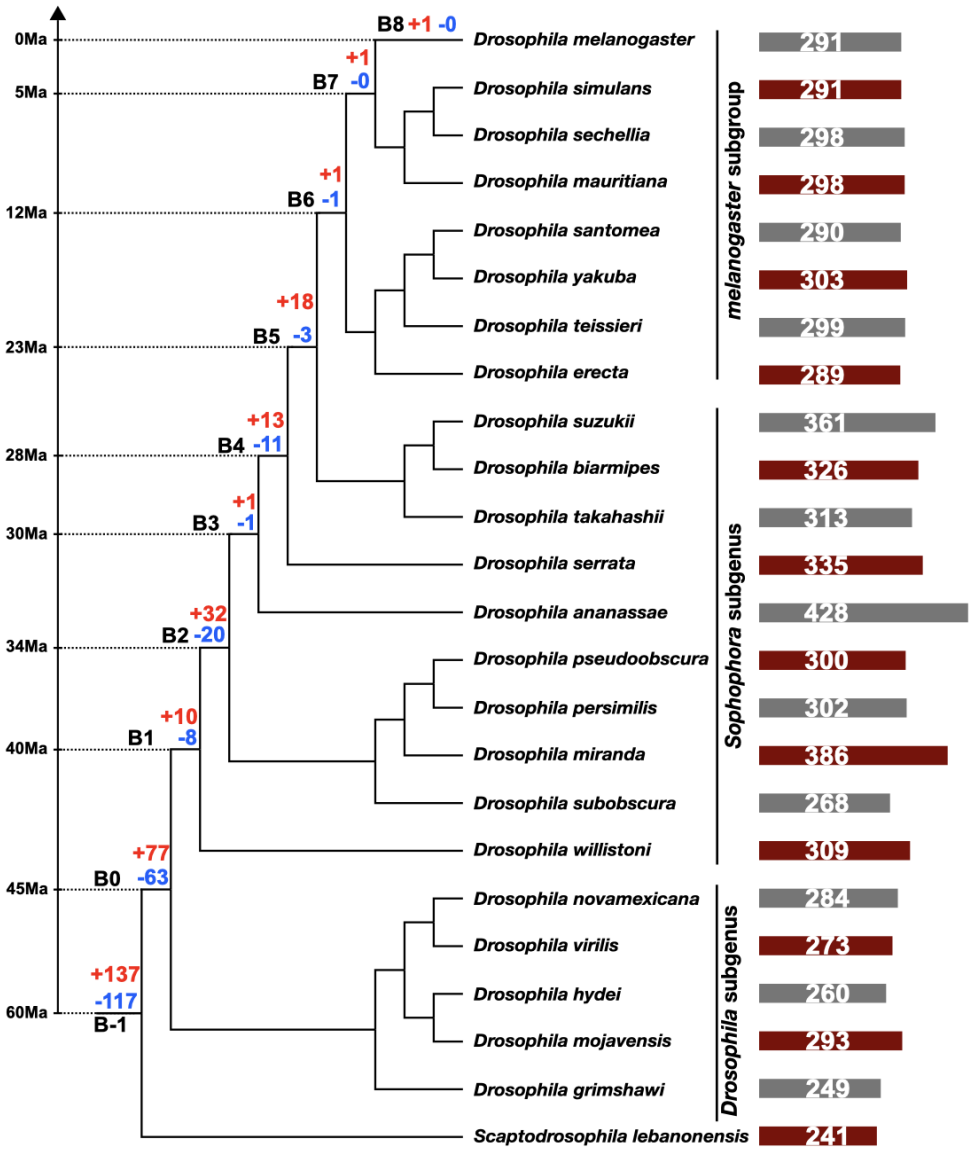



Figure 2.6: Drosophilid species divergence time phylogeny. See references for divergence estimation [Claudia A. M. Russo and Nei, 1995, Obbard et al., 2012]. Values in red indicate the number of *D. melanogaster* tRNA gene births at a given branch while blue values represent the total number of loss events of genes originating from the previous branch.

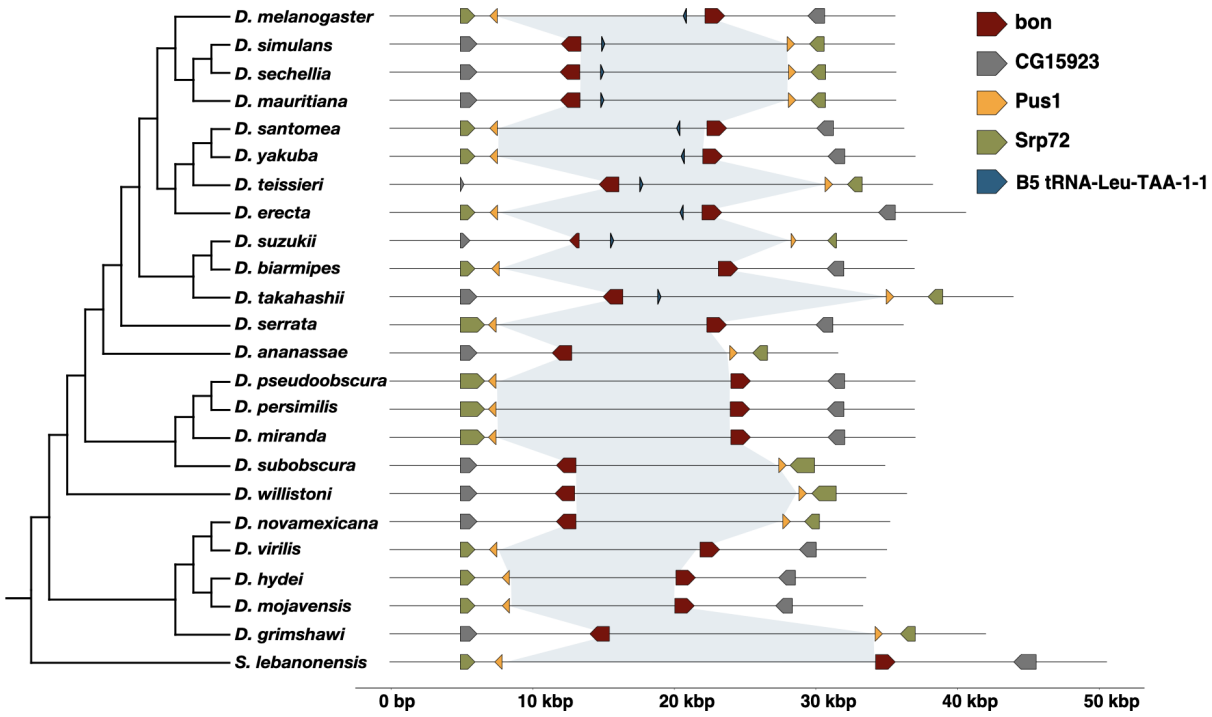



Figure 2.7: Schematic illustrating the gene age dating method applied to Drosophilid genomic tRNAs. For example, we inferred the time of origination of the young isodecoder *tRNA-Leu-TAA-1-1* as occurring at branch 5 (B5), approximately 23Ma ago, as the earliest diverging species containing a copy of this gene is *D. takahashii* in the clade just before the divergence of the *melanogsaster* subgroup. Blue shaded area represents the conserved syntenic area used for homology detection.

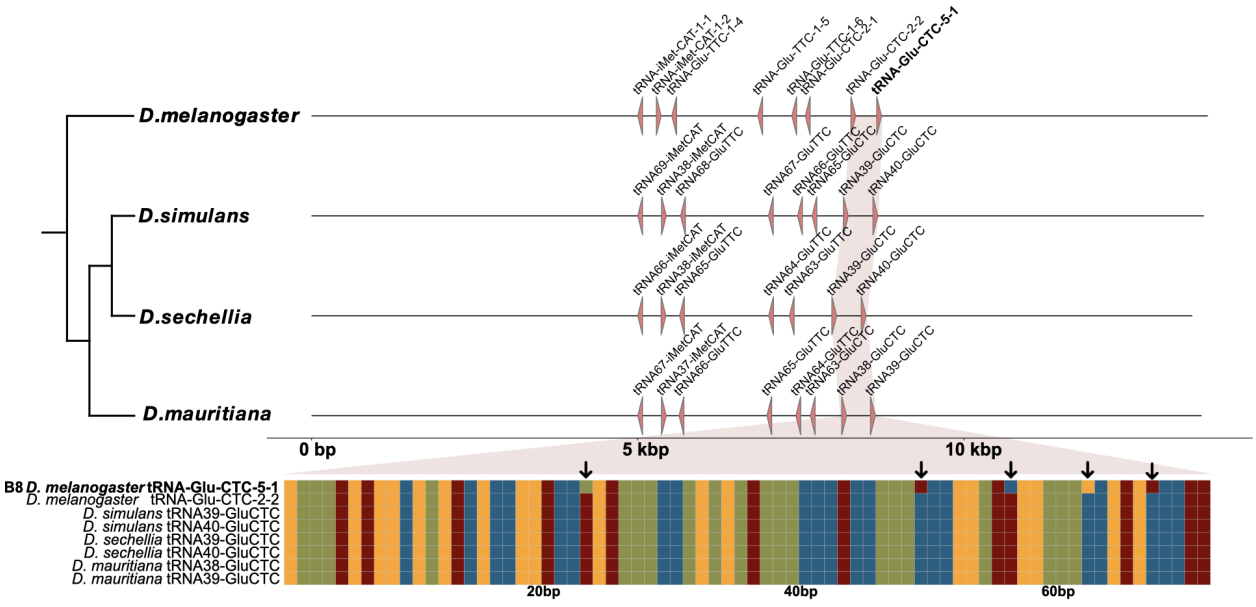



Figure 2.8: Multiple sequence alignment of eight $tRNA^{Glu}$ genes from four species in the *melanogaster* subgroup highlighting the sequence divergence of the *D. melanogaster* specific isodecoder *tRNA-Glu-CTC-5-1*.

The fixation of gene loss is also an important contributor to genomic and phenotypic evolution [Guijarro-Clarke et al., 2020]. For a given *D. melanogaster* tRNA gene we determined the presence or absence of an orthologous, syntenic tRNA in species which diverged after the inferred time of origination of the focal gene. We estimated the rate of tRNA gene loss for genes originating in *Sophophora* subgenus and *melanogaster* subgroup species (B1 – B8) to be 1.1 tRNA genes per Ma (44 genes / 40 Ma). In *D. melanogaster* we observed that 21% (62/291) of the tRNA gene repertoire is not lost in any species following their emergence, possibly constituting a "core" set of tRNA genes as has been proposed previously [Rogers et al., 2010]. The anticodons of those tRNAs correspond to 16 genetically encoded amino acids. This suggests some tRNA duplicates primarily perform their canonical role in translation while others that are lost or gained may be free to acquire "forbidden" substitutions as postulated by Ohno facilitating the acquisition of copy-specific, non-canonical

Table 2.2: Number of tRNA gene gains and losses inferred by gene age dating analysis.

| **Branch** | **Gain** | **Loss** |
|---|---|---|
| B-1 | +137 | -117 |
| B0 | +77 | -63 |
| B1 | +10 | -8 |
| B2 | +32 | -20 |
| B3 | +1 | -1 |
| B4 | +13 | -11 |
| B5 | +18 | -3 |
| B6 | +1 | -1 |
| B7 | +1 | -0 |
| B8 | +1 | -0 |

functions [Ohno, 1970, Zhu et al., 2024]. Of those tRNAs that are retained in all species considered in this work, 32 are unique isodecoding sequences which represent 36% of all *D. melanogaster* isodecoder subfamilies. This suggests increased nucleotide diversity through isodecoder substitutions may allow non-exact duplicates to be maintained over long periods of evolutionary time beyond speciation events. Thus, while mRNA translation is a deeply conserved process tRNA gene repertoires are dynamic and rapidly evolve in copy number and isotype over short evolutionary timescales.

### *2.3.3 Genomic location of tRNA duplicates is related to gene age*

As duplication is the dominant mechanism of tRNA gene origination, we asked which types of duplications (tandem, dispersed) are shaping the *D. melanogaster* tRNA gene repertoire and how new tRNA genes are distributed interchromosomally. Previous studies have reported a general "out of X" trafficking pattern of newly evolved protein coding DNA- and RNA-based duplicates in *D. melanogaster* coupled with sex tissue-specific expression [Vibranovski et al., 2009]. To study the inter- and intrachromosomal movement of duplicated tRNA genes in *D. melanogaster,* we first identified putative parental genes of the *Sophophora*-specific (B1 – B8) genes ($n$ = 175) through local alignment with 1kb flanking regions. Assuming in-

terchromosomal gene movements are proportional to a source chromosome and the ability of a target chromosome to receive a gene insertion is proportional to its size, we computed the expected frequency of duplication directions following [Betran et al., 2002]. After inferring putative duplicate gene pairs, we compared frequencies of different interchromosomal movement events to observed. While the majority of the 75 *Sophophora* subgroup-specific tRNA genes we studied are tandemly duplicated on autosomes (83%), we observed a 943% excess of tandem duplication on chromosome X than expected which was determined to be significant through a multinomial Monte Carlo resampling test with 10,000 simulations ($p = 0.000119$, Table 2.3) [Emerson et al., 2004], in contrast to the gene trafficking hypothesis observed in coding genes. Gene overlap, also termed overprinting, is one mechanism by which prokaryotic and eukaryotic genomes expand their functional potential. We hypothesized that gene overlap could facilitate the emergence of new tRNA genes, particularly if the diversification of tRNA repertoires is a repetitive, selfish process skewed toward tandem duplication [Charlesworth et al., 1994]. We observed that all protein coding genes overlapped by tRNA are ancient, in accordance with hypotheses of selective pressures which overwhelmingly favor loss of new duplicates before their sequences can be selectively modified [Ulfar Bergthorsson and Roth, 2007]. In total we identified 145 tRNA overlapping protein coding gene events, the majority of which are antisense (n=82), spanning tRNA gene ages from B-1 to B5. Surprisingly, a young tRNA we identified as differentially expressed in testis, *tRNA-Asp-GTC-3-1*, overlaps the third intron of *PIG-S,* an ancient subunit of glycosylphosphatidylinositol (GPI) transamidase. We discovered that tRNA overlapping exons can occur on the sense and antisense strands (n=8) of development and gametogenesis related genes such as *non-stop, outspread, polychaetoid, neurotransmitter transporter-like,* and *CG32320* (ortholog of *H. sapiens DCST.*) We also observed nested overlapping structures of several ancient tRNA gene (n=11) clusters of the Lys, Glu, and Tyr isotypes through combination of sense and antisense overlap of two distinct protein coding genes. These results again suggest

the expansion of tRNA repertoires as a repetitive, neutral process that can occur rapidly [Charlesworth et al., 1994].

Table 2.3: Multinomial Monte Carlo resampling test of the difference between observed and expected tRNA gene movement patterns. X → and → X refer to predictions and observations leaving or entering a chromosome, respectively.

| Direction of the gene formation event | Expectation | | Observed | Excess (%) | *P* |
|---|---|---|---|---|---|
| | % | N | N | | |
| X → A | 7.09 | 5.32 | 4 | −24.86 | 0.4918 |
| A → A | 78.96 | 59.22 | 60 | 1.32 | |
| X → X | 1.15 | 0.86 | 9 | 943.63 | 0.0001119 |
| A → X | 12.79 | 9.59 | 2 | −79.15 | |

### *2.3.4 Few recently duplicated isodecoders are directionally selected in natural populations*

Following our observations of tRNA genes originating at various timepoints during *Drosophila* evolution we hypothesized that Darwinian selection would be accelerating the differential emergence and retention of tRNA genes amongst eukaryotic species. We leveraged population genetic data of *D. melanogaster* to test for interpopulation selection signals using the site frequency spectra (SFS) of genomic windows containing tRNA loci. We retrieved 727 accessions of *D. melanogaster* natural population samples representing the putative ancestral range in Africa (AFR) and locations outside of Africa (OOA) [Coughlan et al., 2022]. We computed nucleotide diversity, $\pi$, and the fixation index, $F_{st}$, in 400bp sliding windows with a 50bp step size along the *D. melanogaster* genome. We reasoned that if tRNA genes are continuously duplicating such that interspecific tRNA copy number and isotypes diverge drastically, then they would exhibit significant nucleotide diversity between *D. melanogaster* populations indicative of natural selection. By comparing the ratio of $\pi_{AFR}$ and $\pi_{OOA}$ to

the weighted $F_{st}$ of tRNA loci we identified six candidates: B-1 genes *tRNA-Lys-CTT-1-5*, *tRNA-Gly-TCC-1-1*, *tRNA-Ser-CGA-1-1,* and *tRNA-Trp-CCA-2-6;* B2 *tRNA-Arg-TCG-2-1,* and B4 *tRNA-Lys-CTT-1-11* (Fig. 2.9). Genomic windows containing each of these six tRNA genes had significant reductions of $\pi$ in African populations and increased $F_{st}$, suggesting some isodecoding copies may be adaptively maintained in distinct populations of *D. melanogaster*.

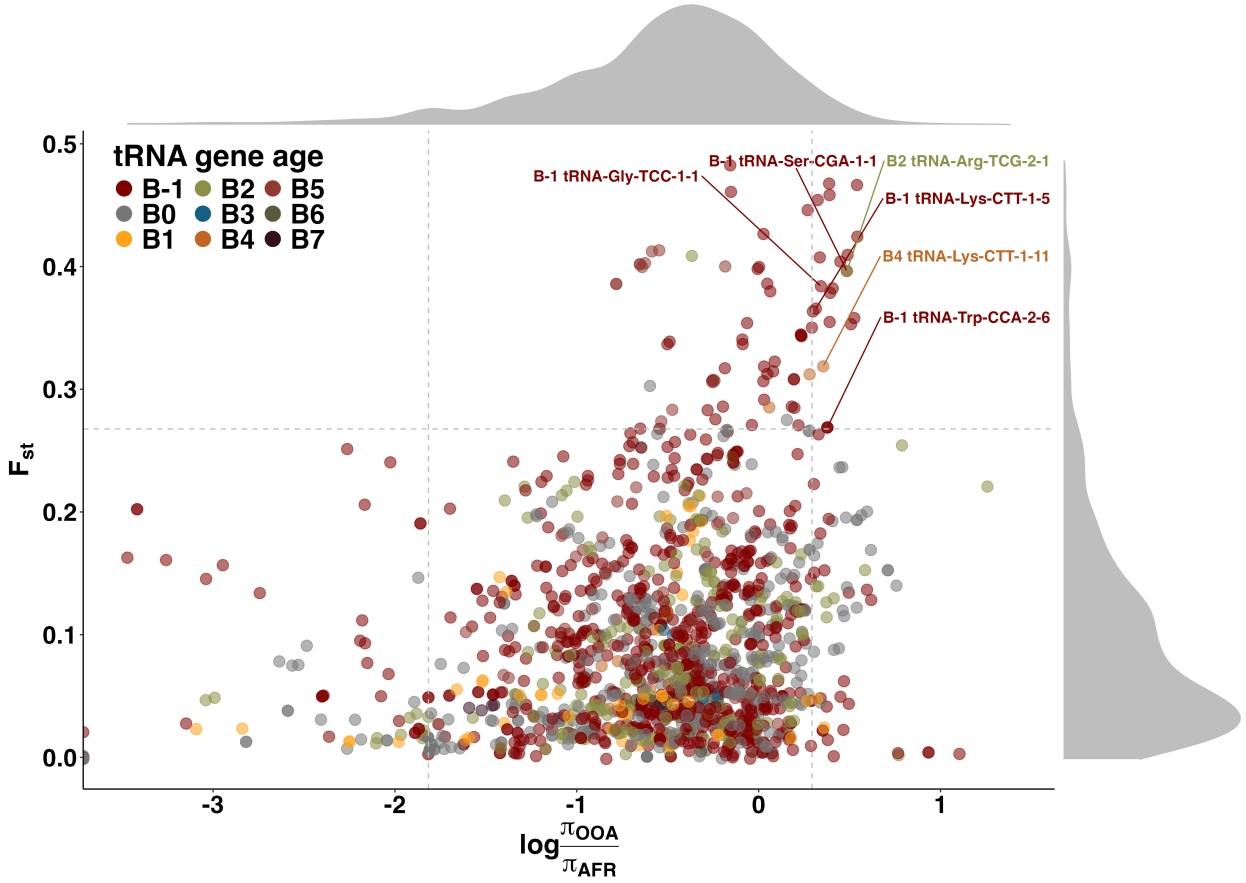



Figure 2.9: Log ratio of tRNA nucleotide diversity ($\pi$) between African (AFR) and outside of Africa (OOA) population samples plotted against the weighted fixation index ($F_{st}$). Gray dashed lines indicate the significance thresholds used ($F_{st}$ $95th = 0.268$, $\pi$ $95th = 0.293$, $\pi$ $5th = -1.82$) for the weighted fixation index ($F_{st}$) and nucleotide diversity ($\pi$) ratio. The six genomic windows containing significant tRNA genes are labeled in the upper right quadrant.

### *2.3.5 Substitutions related to structure, function, and age suggest evolution of functionality*

The paucity of high frequency minor alleles observed at tRNA loci may be attributable to the structure of tRNA which is highly conserved and specialized (Fig. 2.8). We hypothesized that inferred gene age or functionally relevant structures of tRNA may affect SNP frequencies. To this end, we standardized base positions across *D. melanogaster* tRNA genes of varying length by aligning them to a covariance model of eukaryotic tRNA structure with INFERNAL v.1.1 [Holmes AD and TM, 2020, Nawrocki and Eddy, 2013]. SNPs were observed to be present in all structural features of tRNA with the fewest in the anticodon loop and the majority in the acceptor stem (Fig. 2.10), reflecting a functional constraint on SNP location. Loop regions contained fewer SNPs than stems highlighting their functional importance for aminoacylation as well as ribosome and aaRS recognition, suggesting purifying selection against SNPs within loops specifically [Li et al., 2016]. We next tested whether gene age correlated with SNP frequency and location (Fig. 2.11). We observed no significant relationships ($p > 0.05$, Pearson's correlation) between inferred gene age and mean number of SNPs in each of the structural features of tRNAs (Fig. 2.12). These results indicate that tRNA copies originating at different time points experience equivalent purifying selection to reduce deleterious mutations.

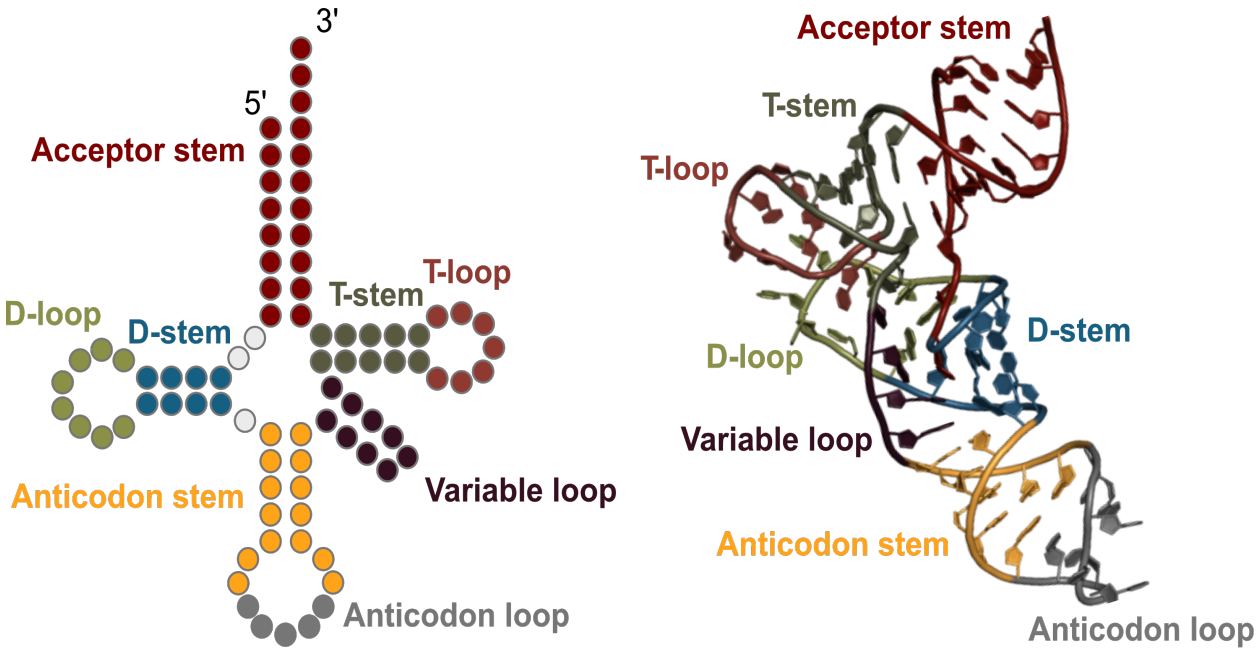



Figure 2.10: (Left) Illustration of tRNA secondary structure. (Right) Predicted tertiary structure of *tRNA-Glu-CTC-5-1*

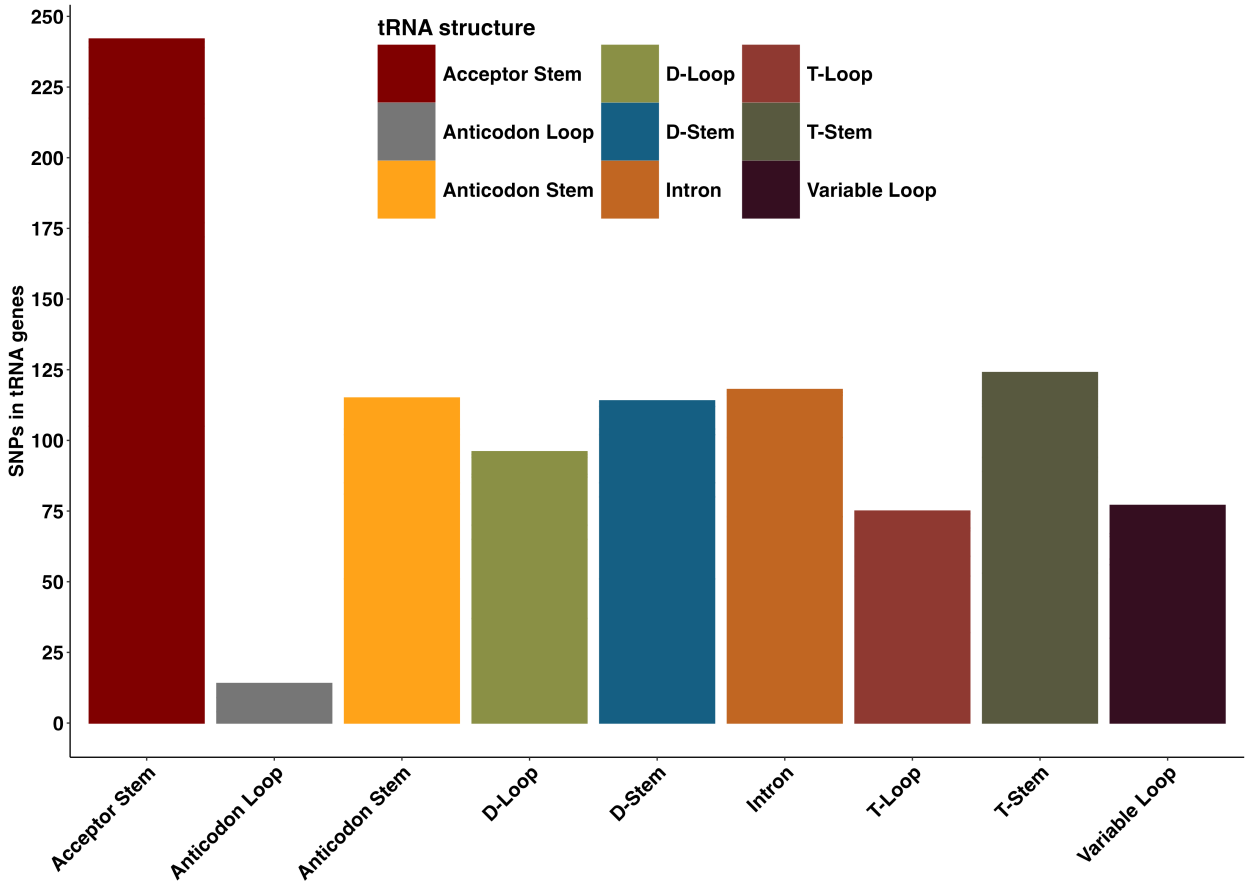



Figure 2.11: Distribution of SNPs within structural features of tRNA.

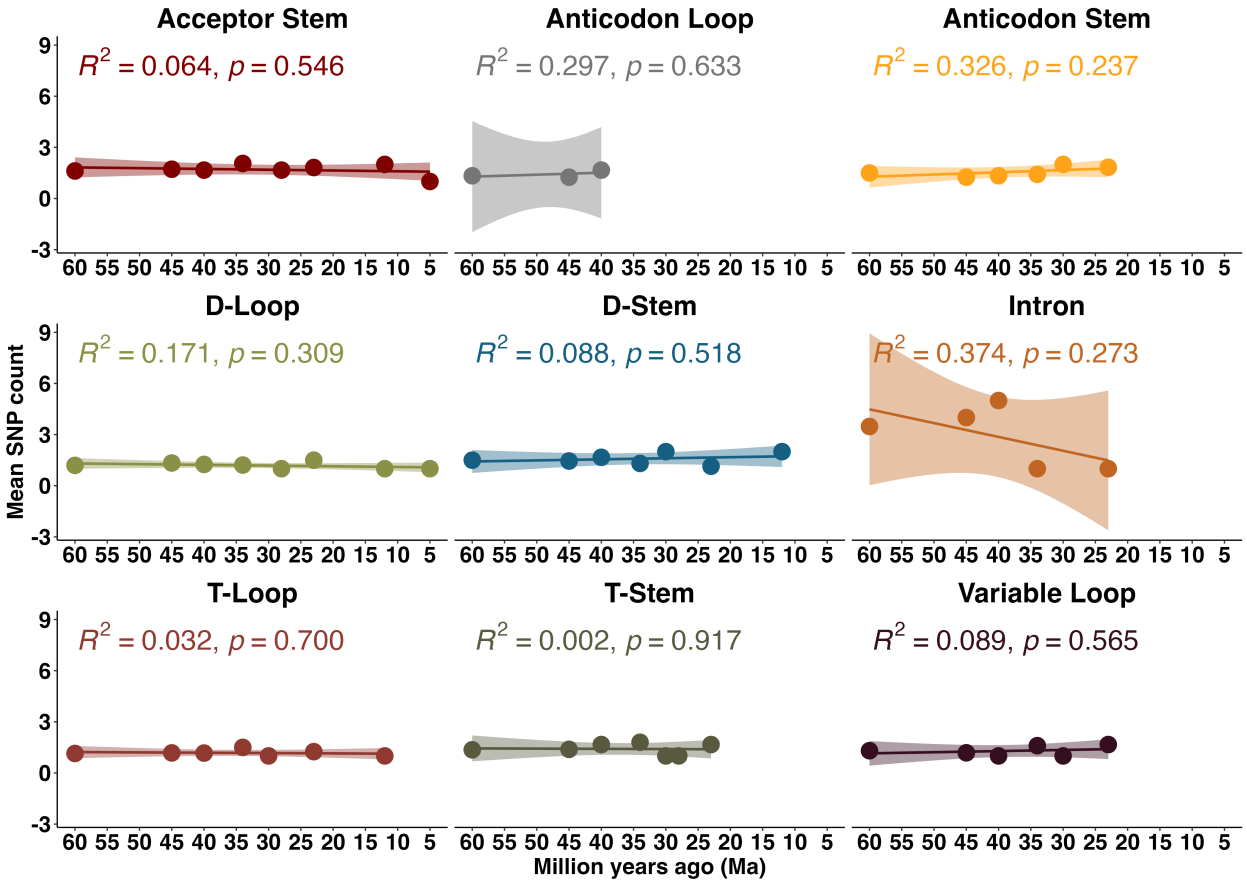



Figure 2.12: Linear regression analyses illustrating the relationship between mean SNP density within structural features and the inferred evolutionary age of tRNAs given in million years (Ma).

Following our observation of similar mutation patterns between tRNA duplicates, we next asked how fixed substitutions in taxon-specific tRNA genes may affect structural integrity and stability. Despite the recent inferred origination of the *melanogaster*-specific isodecoder identified previously, *tRNA-Glu-CTC-5-1,* it has five fixed substitutions compared to syntenic orthologs (Fig. 2.7, Fig. 2.13A,B) and we estimated a substitution rate of $0.0597 bp^{-1} Ma^{-1}$ in this gene using orthologous tRNAs from species spanning B3-B8 [Yang, 2007]. We aligned the predicted structures of *tRNA-Glu-CTC-5-1* to the nearest conserved Glu isoacceptor, B4 gene *tRNA-Glu-CTC-2-2.* The transversion within the taxon-specific *tRNA-Glu-CTC-5-1* at residue C24 disrupted Watson-Crick base pairing of A24:U11, one

of several identity determinants for glutamylation of $tRNA^{Glu}$, potentially allowing a tertiary bond between residues 9 and 10 in the D-loop [Quigley and Rich, 1976, Sekine et al., 1996] (Fig. 2.13C,D). Two substitutions at positions 50 and 63 permitted Watson-Crick base pairing of 50U:63A, eliminating compensatory Hoogsteen interactions of 50G:62C and 63G:49C observed in *tRNA-Glu-CTC-2-2* [Targanski and Cherkasova, 2008], (Fig. 2.13E). The substitution at 57G in the U-turn motif of the T-loop may stabilize tertiary interactions between the D- and T-loops [Jucker and Pardi, 1995], and the orientation of the transition at residue 68A at the 3' end of the acceptor stem is flipped compared to the conserved 68G, potentially disrupting base pairing with 4C (Fig. 2.13F). These data illustrate that despite strong purifying selection and largely a lack of signatures of directional selection, recently duplicated tRNAs can rapidly accumulate substitutions yielding novel isodecoders.

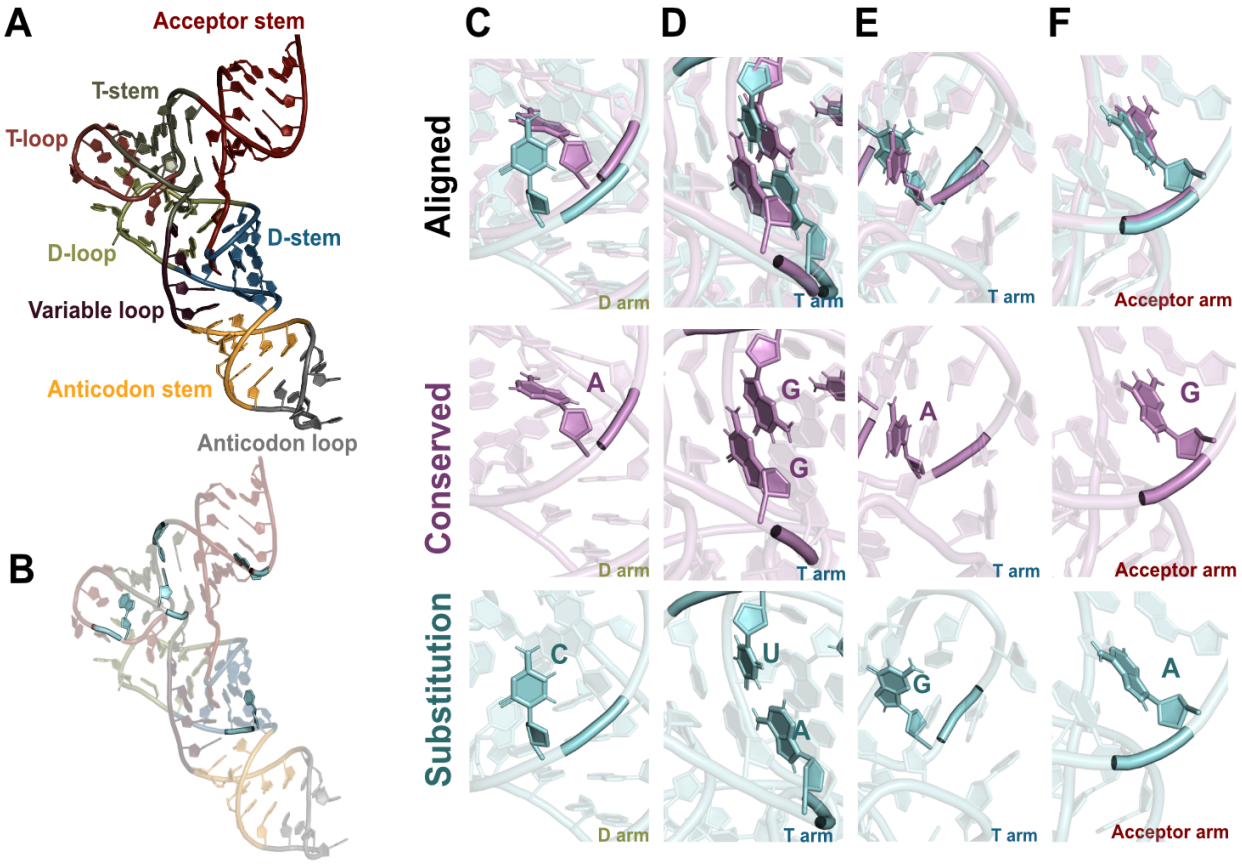



Figure 2.13: (A) Predicted tertiary structure of *tRNA-Glu-CTC-5-1* labeled with structural features. (B) Location of substitutions (blue) within *tRNA-Glu-CTC-5-1*. (C-F) Structural alignment (top row) of predicted tertiary structures of *tRNA-Glu-CTC-2-2* (purple, second row) and the *D. melanogaster*-specific isodecoder *tRNA-Glu-CTC-5-1* (blue, third row).

## 2.4 Materials & Methods

### *2.4.1 Biased synonymous codon usage and tRNA adaptation test*

Coding sequence annotations for *D. melanogaster* were retrieved from NCBI release 6.57. Frequencies of codon usage per gene were computed with codonw v.1.4.4 [Peden, 2005]. Gene-wide biased codon usage was compared using the translational adaptation index (tAI) as described in [dos Reis et al., 2004]. Statistical analysis and visualization of results were performed with custom R scripts [Wickham, 2016].

Briefly, $W_i$ represents the absolute adaptiveness value given as the sum of selective con-

straints on codon-anticodon coupling $s_{i,j}$ multiplied by tRNA gene copy number, $tGCN_{ij}$, for the $j_{th}$ tRNA that recognizes the $i_{th}$ codon to the $n_{th}$ isoacceptor that recognizes the $i_{th}$ codon, excluding stops. Relative adaptiveness values of codons are given by $w_i$ given by $W_i/W_{\max}$ if the codon's absolute adaptiveness is not equal to 0, or the geometric mean of all $w_i$ with $W_i \neq 0$ otherwise. Adaptation of a gene to the genomic tRNA pool is represented by $tAI_g$ as the geometric mean of the relative adaptiveness values of $k$ codons along the length $l$ of a gene $g$. Effective number of codons $N_c$ of a gene $g$ were computed using an optimized version of Wright's equation, $f\left(x_g\right)$, [Wright, 1990] as described in [dos Reis et al., 2004]. Estimates of translational selection for biased synonymous codon usage was quantified as the difference between the expected $N_c$ under neutrality and observed $Nc_g$. The non-parametric correlation $S$ represents co-adaption between $\psi_g$ and $tAI_g$ which is a single value between -1 and 1 where high correlation coefficients indicate stronger translational selection.

### *2.4.2 Genomic sequence retrieval and annotation*

24 Drosophilid long-read assemblies and their protein coding gene annotations were downloaded from NCBI. For each of the 24 assemblies we annotated tRNA genes using tRNAscan-SE v.2.0.9 [Chan et al., 2021], discarded predicted pseudogenes, and kept only a set of high confidence annotations for downstream analyses.

### *2.4.3 Gene age dating*

A Drosophilid 24-species divergence time estimated phylogeny was obtained from TimeTree [Kumar et al., 2017]. For each of the 291 *D. melanogaster* tRNA genes we first identified its two nearest protein coding genes flanking each side to create a set of four protein coding "anchor" genes. These anchor genes delineated a region containing one or more tRNA genes flanked by the anchors. By repeating this procedure, we identified tRNA and anchor gene sets in the 23 outgroup Drosophilid species which together with the *D. melanogaster*

data were used to search for orthologous tRNA genes without convolution by extraneous, non-syntenic sequences. Our method largely reduced the homology search space and the probability of incorrectly aligning tRNA genes which would lead to over- or underestimation of gene ages by determining syntenic regions between conserved protein coding genes. For each *D. melanogaster* tRNA and anchor gene set, we identified orthologs and extracted their genomic region between their two anchor genes. We then conducted local alignments using the extracted outgroup anchor region and a given D. melanogaster tRNA gene to determine ortholog relationships [Camacho et al., 2009]. With the resulting alignments D. melanogaster tRNA gene ages were assigned according to a parsimony-based strategy similar to the methodologies of Zhang et. al. [Zhang et al., 2010a] and Zhou et. al. [Zhou et al., 2008]. Briefly, the age of a *D. melanogaster* tRNA gene was inferred based on the presence or absence of syntenic orthologs in the other 23 Drosophilid species. If a tRNA gene had orthologs in multiple species, the gene age was assigned to the phylogenetic branch of the oldest species with an alignment length of at least 70bp to avoid underestimation of gene ages. We further conducted a branch-wise analysis of tRNA gene presence and absence to identify tRNA loss events throughout the Drosophila phylogeny. If a gene was present in species on branch X but lacks a syntenic ortholog in any of the species on the next branch or younger, the gene would be inferred as lost in a given species.

### *2.4.4 Chromosomal movement*

We extended the methodologies of Vibranovski et. al. and Betrán et. al. [Betran et al., 2002, Vibranovski et al., 2009] to compute the expected frequency of tRNA duplications between and within *D. melanogaster* chromosomes. We computed reciprocal local alignments of the 77 *D. melanogaster* tRNA genes we annotated as emerging since the divergence of the Sophophora subgenus (B1) with a 1kb flanking region on either side [Quinlan and Hall, 2010]. From the resulting alignments we determined putative parental genes of younger

tRNA genes, and their type of duplication (tandem, dispersed) based on hits with minimal mismatches, gaps, and e-value, and maximum length of alignment to the query.

### *2.4.5 Whole-genome variant calling*

Natural populations of *D. melanogaster* with whole genome short read resequencing (WGS) data were downloaded from NCBI SRA [Coughlan et al., 2022, Rogers et al., 2014]. In total, we obtained 727 high-quality *D. melanogaster* WGS samples after filtering data with a sequencing coverage threshold of at least 10. The WGS mapping was based on the standard pipeline of GATKv4.1 [Poplin et al., 2018]. Briefly, we first conducted quality control, adapter trimming, and quality filtering for raw reads with FastQC v2 [de Sena Brandine and Smith, 2021] and fastp [Chen et al., 2018]. The cleaned reads for each sample were mapped to the dm6 reference genome using BWA v0.7.17 [Li and Durbin, 2009] with default parameters. Following the sorting, indexing, and marking of PCR duplicates, the variants were called with the pipeline of the Best Practice protocol of GATK v4.1 using HaplotypeCaller [Poplin et al., 2018]. All variants were then filtered based on the following threshold: quality score of 30, sequencing coverage of 10x, genotype quality of 30, and a maximum of 25% missing data [Coughlan et al., 2022]. The allele frequencies were analyzed with PLINK v1.9 [Purcell et al., 2007]. These data were used for downstream polymorphism analyses.

### *2.4.6 Structural analysis*

Tertiary tRNA structures were predicted with trRosettaRNA [Zhang et al., 2022] and structurally aligned with US-align [Wang et al., 2023a]. Visualization was performed with PyMol [Schrödinger, 2015].

# CHAPTER III

# SMALL RNA FUNCTION IN RELATION TO SEX

## 3.1 Abstract

Previous conceptions of tRNA gene evolution have largely focused on the molecule's central role in mRNA translation and its relationship to codon composition. We show that sexually dimorphic functions of individual tRNA copies have been underappreciated in molecular and evolutionary considerations of tRNA complements. Through multiplexed small RNA sequencing of ovary and testis tissues, we identified sex-biased expression in 38 tRNA genes across various gene ages, primarily as antisense or precursor transcripts, suggesting non-translational functions like signaling or regulation are influencing the generation and maintenance of tRNA duplicates. Notably, the youngest differentially expressed tRNAs, *tRNA-Lys-TTT-2-5* (upregulated in ovaries) and *tRNA-Asp-GTC-3-1* (upregulated in testes), highlight rapid functional divergence of these essential non-coding RNAs. Additionally, 43 isodecoders showed sex-specific tRNA fragmentation patterns, underscoring the importance of sex in tRNA functional diversification beyond decoding efficiency. We further test for signatures of adaptive evolution by integrating population genetic data from *D. yakuba* and *D. simulans* to compute statistical tests of neutrality.

## 3.2 Introduction

Beyond their primary role as mRNA decoders, tRNAs accomplish numerous additional non-canonical functions. These include: acting as regulatory small RNAs through tRNA-derived fragments (tRFs and tiRNAs), which modulate gene expression, translational control, stress responses, intergenerational inheritance, apoptosis inhibition via cytochrome c binding, retrotransposon silencing, and antiviral defense [Su et al., 2020, Akiyama and Ivanov,

2023, Keam and Hutvagner, 2015, Goodarzi et al., 2015, Schorn and et al., 2017, Mei et al., 2010, Chen et al., 2016, Sharma et al., 2016, Yamasaki et al., 2009]. Additionally, tRNAs participate in chromatin insulation, mRNA decay recruitment, and innate immune sensing [Raab et al., 2012, Zhu et al., 2024, Schlee and Hartmann, 2016]. Further, specific tRNA fragments have been implicated in tissue repair, cancer progression, antiviral defense, and even the transmission of paternally acquired metabolic phenotypes through sperm [Goodarzi et al., 2016, Li and et al., 2018, Chen et al., 2016, Zhang et al., 2018].

These non-canonical activities are often modulated by tRNA modifications, tissue-specific expression patterns, and dynamic changes in tRNA pools, highlighting the multifaceted regulatory potential of tRNAs in eukaryotic biology [Pan, 2018, Machnicka et al., 2014, Pinkard et al., 2020, Orellana et al., 2022]. In multicellular organisms, tRNA expression varies across tissues, reflecting specialized translational demands [Dittmar et al., 2006, Pinkard et al., 2020]. Despite tRNA's centrality to all life, evolutionary studies of tRNA have directed limited attention to quantitative tRNA sequencing. Prediction methods like tRNAscan-SE have greatly advanced tRNA gene annotation, expanding comparative genomic potential, but functional expression profiling has been lagged, particularly in capturing post-transcriptional modifications and processing that affect tRNA stability and function [Chan et al., 2021, Chan and Lowe, 2015, Pan, 2018]. Without those data, limited conclusions can be drawn about the evolutionary dynamics of tRNA gene families, the selective pressures shaping their diversity, and the functional consequences of tRNA variation across tissues and species. In this chapter, we utilize a recently developed tRNA sequencing method to quantify the abundance, fragmentation, and modification of individual tRNA genes in *D. melanogaster* gonads for the first time [Watkins et al., 2022]. We then investigate selective pressures affecting individual tRNA genes, with particular focus on those displaying differential expression.

## 3.3 Results & Discussion

### *3.3.1 Lineage-specific tRNA duplicates demonstrate sex-biased expression and fragmentation*

As sexual and sexually antagonistic selection have been found to bei driving forces in the emergence of new protein coding genes, we hypothesized that sex-related selection could underlie tRNA duplication or differences in their abundance or modification [Harrison et al., 2015]. Initially, we planned to perform comparative analyses of interspecies differential expression between seven Drosophilid species. In total, we dissected and froze head, whole body, and gonads samples in biological triplicate from six Drosophilid species: *D. melanogaster, D. simulans, D. yakuba, D. erecta, D. virilis, D. pseudoobscura,* and *D. willistoni* (Table 3.1). However, due to funding limitations we were only able to complete sequencing of *D. melanogaster*. To comprehensively interrogate whether sex or gene age affect the expression of tRNAs, we dissected gonad tissues (ovary, testis) from Iso-1 *D. melanogaster* adults in biological triplicate and performed multiplexed small RNA sequencing of total RNA extractions (MSR-seq) [Watkins et al., 2022]. We observed sexually dimorphic abundances of isodecoders and individual isoaccepting duplicates that originated at different times along the *Drosophila* phylogeny (3.1A,B). One of the youngest tRNA isodecoders we identified previously, *tRNA-Lys-TTT-2-5* which duplicated in the *melanogaster* subgroup at B7 just prior to the divergence of *D. melanogaster*, was significantly upregulated in the ovary (Fig. 3.1C) as a precursor tRNA transcript (pre-tRNA, $p < 0.05$) suggesting sex-specific, non-translational functions [Akiyama and Ivanov, 2023]. Moreover, the next youngest differentially expressed tRNA, *tRNA-Asp-GTC-3-1* which we inferred to have originated in the *Sophophora* subgenus at B5 prior to the divergence of the *melanogaster* subgroup, is upregulated in the testis as an antisense transcript ($p < 0.001$) and may perform male-specific signaling or transcriptional regulatory functions through novel interactions with mRNAs [Luo et al., 2018]. In total,

we identified 38 tRNA genes spanning B-1, B0, B1, B2, B5, and B7 that are differentially expressed between the sexes. These upregulated tRNA transcripts represent 10 isoaccepting families and are primarily antisense or pre-tRNA, suggesting sexual selection not of mRNA decoding efficiency, but of extra-translational functions of tRNAs or their fragments sufficient to maintain redundant isoacceptors and isodecoders in the shared genome.

Table 3.1: List of Drosophilid species dissected. For each species we collected gonads, whole bodies, and heads from male and female adults in biological triplicate.

| **Species** | **ID** |
|---|---|
| *D. yakuba* | 14021-0261.01 |
| *D. virillis* | 15010-1051.87 |
| *D. erecta* | 14021-0224.01 |
| *D. melanogaster* | iso-1; 14021-0231.36 |
| *D. willistioni* | 14030-0811.21 |
| *D. pseudoobscura* | 14011-0121.94 |
| *D. simulans* | RgSim5 |

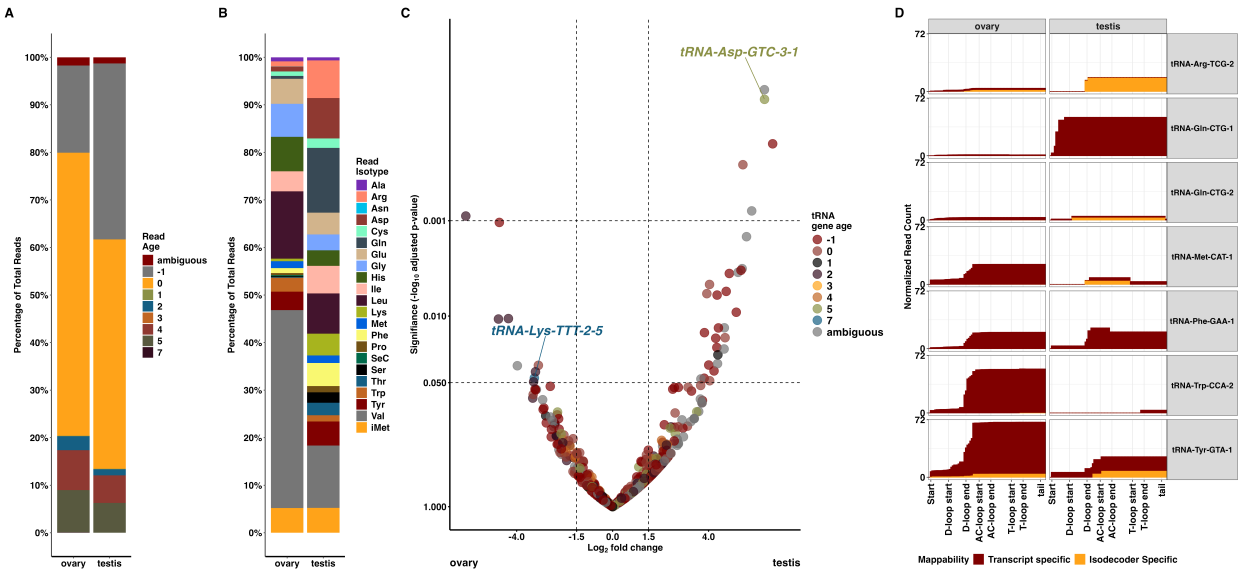



Figure 3.1: The *D. melanogaster* tRNA repertoire exhibits sex-specific expression, fragmentation, and modification. (A and B) tRNA sequencing read distributions between *D. melanogaster* gonads. (A) Proportion of tRNA-seq reads colored by inferred gene age and (B) distribution of reads colored by tRNA read isotype. (C) Volcano plot of differentially expressed tRNAs genes between ovary and testis (rlog normalization [Holmes AD and TM, 2020]). Points are colored by gene age unless there is ambiguity due to identical tRNA sequence. (D) Bar plots showing the start and end positions of ovary- and testis-specific patterns of tRNA fragment abundance colored by specificity of reads. Red bars indicate reads specific to an isotype; yellow bars are isodecoder specific reads.

In addition to abundance, we asked whether tRNA fragmentation, an increasingly appreciated source of tRNA functional diversity is also variable between sexes [Randau and Soll, 2008]. We observed tissue- and isodecoder-specific differences in fragmentation patterns by determining the start and end position of fragmented transcripts (Fig. 3.1D). Overall, we observed 43 isodecoders as the source of sexually dimorphic fragments (Data S4) between *D. melanogaster* reproductive tissues. Identical tRNAs exhibited sex-specific fragmentation patterns such as *tRNA-Trp-CCA-2,* which is highly upregulated in ovary, but lowly expressed as a 3' fragment in testis; *tRNA-Gln-CTG-1* which is upregulated in testis but not ovary, *tRNA-Tyr-GTA-1* which is expressed in both tissues, but is biased toward the ovary; and *tRNA-Phe-GAA-1* isoacceptors which are equivalently expressed in both sexes, save for a fragment consisting of the D-loop end through the anticodon loop end. Further,

we identified instances of isodecoder and sex-specific fragmentation such as *tRNA-Met-CAT-1, tRNA-Arg-TCG-2,* or *tRNA-Gln-CTG-2* which have testis-limited expression exclusively from those isodecoding sequences and not others of the same isotype, underscoring the functional divergence of nearly identical isodecoding tRNAs between reproductive tissues and highlighting the importance of sex in understanding tRNA duplication.

As post-transcriptional modification of tRNAs is essential for their maturation, function, regulation, and conformational rigidity [Novoa et al., 2012, Pan, 2018], we asked if sex or gene age are associated with certain modifications. We tested this by characterizing the distribution of base modifications along the lengths of tRNA transcripts, and we focused our analyses on modified nucleosides found exclusively in either sex which we observed at positions: 14–27,32, 34,37,58, and e9 (Fig. 3.2A). Age distributions of tRNAs with sex-specific modifications were similar between the ovary and testis, but ovary-specific base modifications occurred primarily on ancient tRNAs (Fig. 3.2B,C). Most sex-specific modifications identified were female-biased, suggesting unique tRNA remodeling is necessary to meet translational or extra-translational requirements of the ovary. Strikingly, we observed that young isodecoding genes of that same isotype such as *tRNA-Trp-CCA-2* and *tRNA-Trp-CCA-1*, which originated at B5 after the divergence of the *melanogaster* subgroup, exhibit sexually dimorphic post-transcriptional modification at G26, likely $N^2$-methylguanosine ($m^2G$) and $N^2$,$N^2$-methylguanosine ($m^2{}_2G$), respectively [Machnicka et al., 2014].

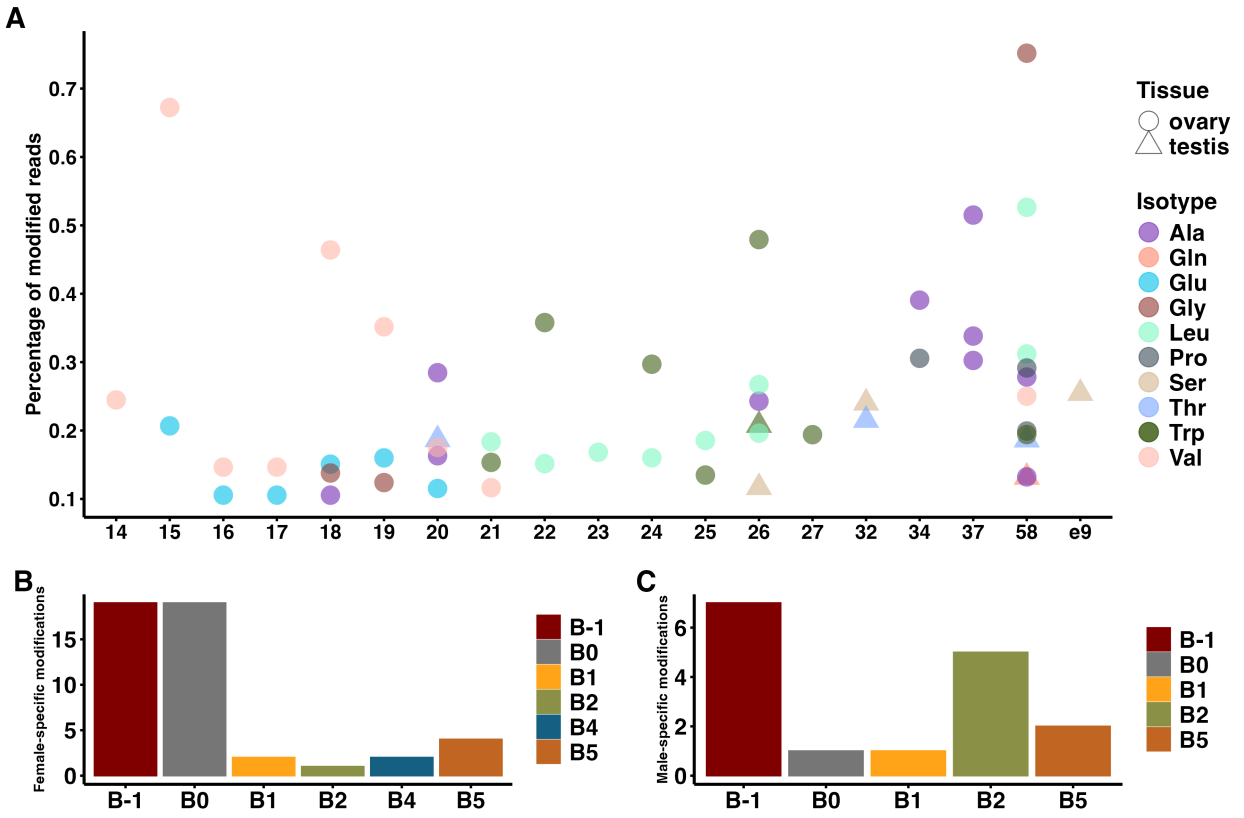



Figure 3.2: (A) Scatter plot of post-transcriptional modifications specific to either sex. Circles indicate ovary-specific modified transcripts while triangles indicated testis-specific modified transcripts. Colors indicate tRNA isotype. (B and C) Bar plots illustrating counts of sex-specific post-transcriptional modifications colored by gene age. (B) Ovary-specific modifications. (C) Testis-specific modifications.

To investigate whether tRNA tissue-specific expression is generally correlated with gene age, as has been reported for recently evolved protein coding genes [Vinckenbosch et al., 2006, Zhang and Zhou, 2019], we computed the tissue specificity measure (spm) [Camargo, 2020] for all tRNA transcripts. When discriminating transcripts by isotype we found that antisense transcripts across isoacceptor groups were biased toward the testis, again suggesting inhibitory functions of tRNA derived fragments necessary to modulate the transcriptomic complexity of the testis which is uncorrelated with other tissues across metazoans (Fig. 3.3) [Innocenti and Morrow, 2010, Soumillon et al., 2013]. The opposite was observed for whole tRNA transcripts which were observed to be primarily specific to the ovaries. These data sug-

gest gene age is related to the rate of tRNA base modification and their extra-translational functions, possibly promoting sex-limited neofunctionalization of new duplicates or their fragments. Thus, selection of sex-biased expression may also underlie tRNA repertoire evolution and their functional diversification.

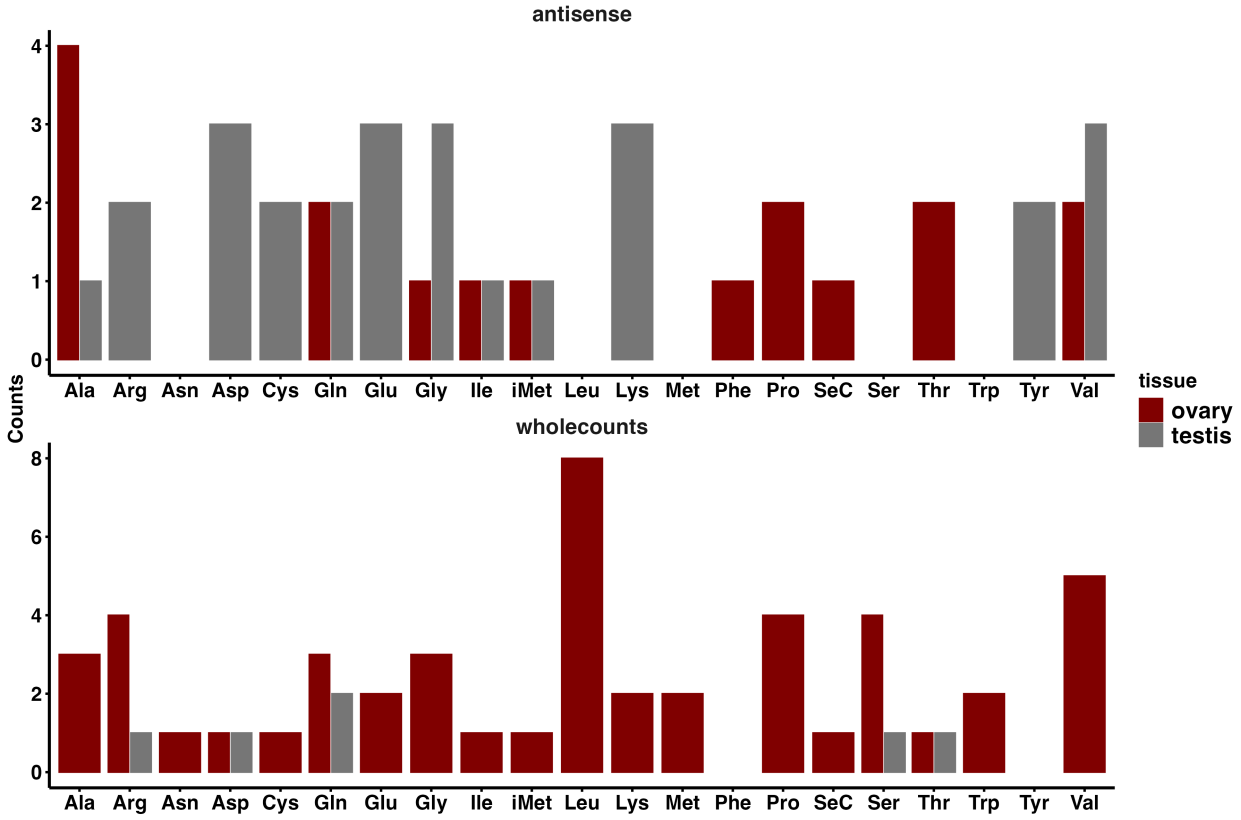



Figure 3.3: Bar plots of tRNA-seq reads with biased tissue specificity ($spm \geq 0.9$), where red indicates ovary and gray indicates testis (antisense above and whole tRNA below).

### *3.3.2 Some tRNAs may be selected for sex-associated functions*

By incorporating natural population data of *D. yakuba* and *D. simulans* we polarized the SFS of each tRNA gene with 20kb flanking regions to test for selective sweeps at tRNA loci through statistical tests of neutrality (Fig. 3.4). We generated null SFS distributions by the method of Hudson [Hudson, 2002] with 10,000 simulations using a *D. melanogaster* population scaled mutation rate $\theta = 0.01328$ [Hudson et al., 1987, Wang et al., 2023b]

and sought to identify departures from neutral expectations by applying Tajima's D and Fay-Wu's H tests to the polarized SFS. For all window and step configurations tested we were unable to reject the null hypothesis ($p > 0.05$) of the D and H tests, indicating non-directional variations in allele frequencies at tRNA loci. As the strongly conserved structure and short sequence of tRNAs may limit the power of spectra-based tests, we next conducted a Hudson-Kreitman-Aguadé-like (HKA-like) test to assess whether tRNA loci exhibited inter- and intra-specific site variability consistent with natural selection [Ford and Aquadro, 1996]. We observed 36 tRNA genes with significant ratios of polymorphic to interspecific divergent sites compared to neutral expectations across all window and step configurations considered. These represented 17 isotypes and gene ages ranging from B-1 to B7 (Fig. 4B, C). In total, four tRNA genes were both differentially expressed in gonads and had significant ($p \leq 0.05$) HKA ratios. Strikingly, two of those four were *tRNA-Lys-TTT-2-5* and *tRNA-Asp-GTC-3-1* which we previously identified as the youngest differentially expressed tRNAs in *D. melanogaster* gonads (Fig. 3.5). Further, when scrutinizing proportions of codon usage in *D. melanogaster* considering our differential expression and population genetic analyses, we found that *tRNA-Lys-TTT-2-5* and *tRNA-Asp-GTC-3-1* are the only two isoacceptors which correspond to the preferred codon of their isotype, have signatures of selection, and are differentially expressed between reproductive tissues (Fig. 3.6). Despite these findings, TTT is not the most abundant anticodon of $tRNA^{Lys}$ isoacceptors by copy number. Of the remaining 36 tRNA genes with signatures of selection, six correspond to a major codon while four conserved isodecoders: B0 gene *tRNA-Ala-AGC-2-11* and B-1 genes *tRNA-Ala-AGC-2-10, tRNA-Ala-AGC-2-9, tRNA-Ala-AGC-2-8* are also differentially expressed. While we observed sexually dimorphic fragmentation patterns of tRNAs from 32 isoaccepting groups, 13 were fragments of tRNA genes at genomic regions under directional selection, and only six in testis and 12 in ovary were derived from tRNAs corresponding to the preferred codon of their isotype (Fig. 3.7). Most of the *D. melanogaster* tRNA repertoire (221/291) is

neither differentially expressed between male and female reproductive tissues nor within genomic regions associated with non-neutral allele frequency variation. These data suggest tRNA repertoire composition can rapidly fluctuate while certain copies are maintained in metazoan populations through selection of sex-associated function.

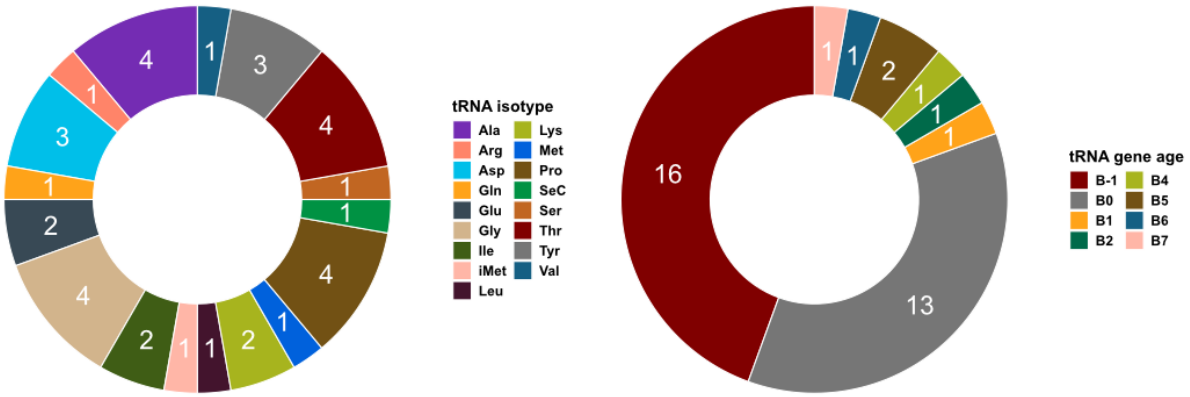



Figure 3.4: (B and C) 36 tRNA genes are within genomic regions with significant HKA-like tests results ($p<0.05$) as determined through simulation tests[Hudson et al., 1987, Hudson, 2002]. (B) 17 isotypes are represented by the naturally selected tRNA genes. (C) Gene age distribution for the 36 significant tRNA genes. (D) Two recently duplicated tRNAs (B7 and B5) *tRNA-Lys-TTT-2-5* and *tRNA-Asp-GTC-3-1* are in regions with significant reductions in proportions of segregating to divergent sites polarized with *D. simulans* and *D. yakuba* as compared to neutral expectations (shown in dark blue is the tRNA gene locus, light blue is the 1000bp window centered on the gene).

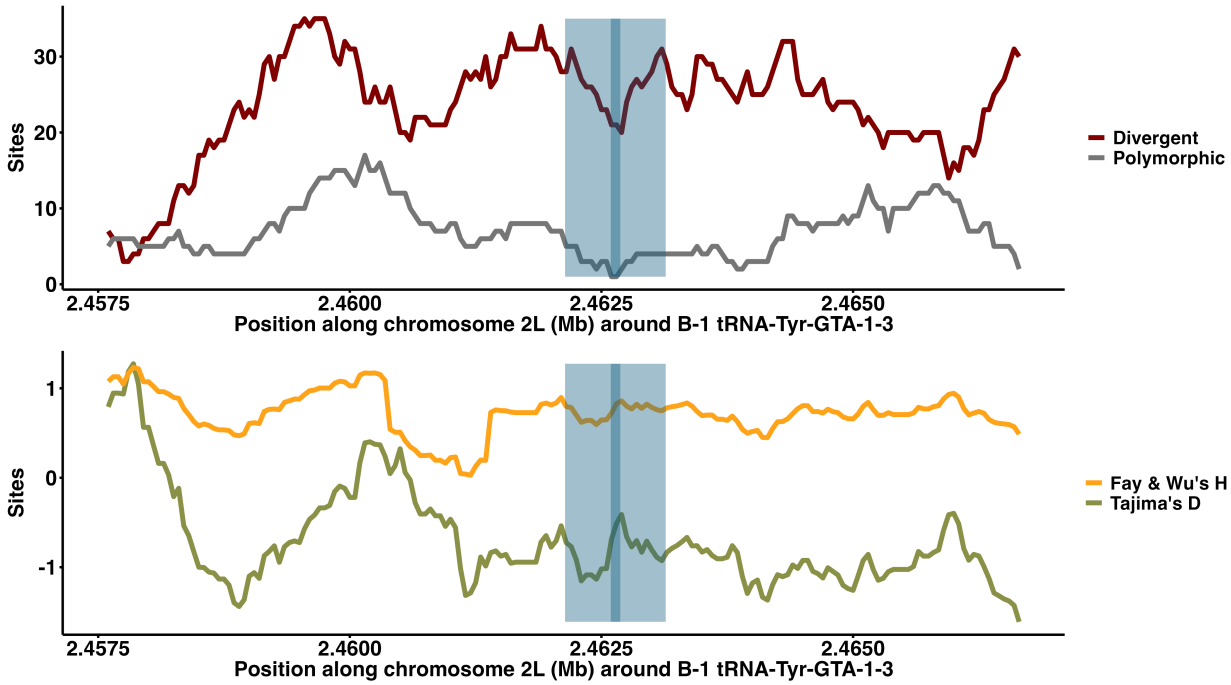



Figure 3.5: Example of SFS-based tests of neutrality for *tRNA-Tyr-GTA-1* for illustration purposes. The SFS of the tRNA gene and 1000bp flanking region are polarized with *D. simulans* and *D. yakuba* as outgroups.

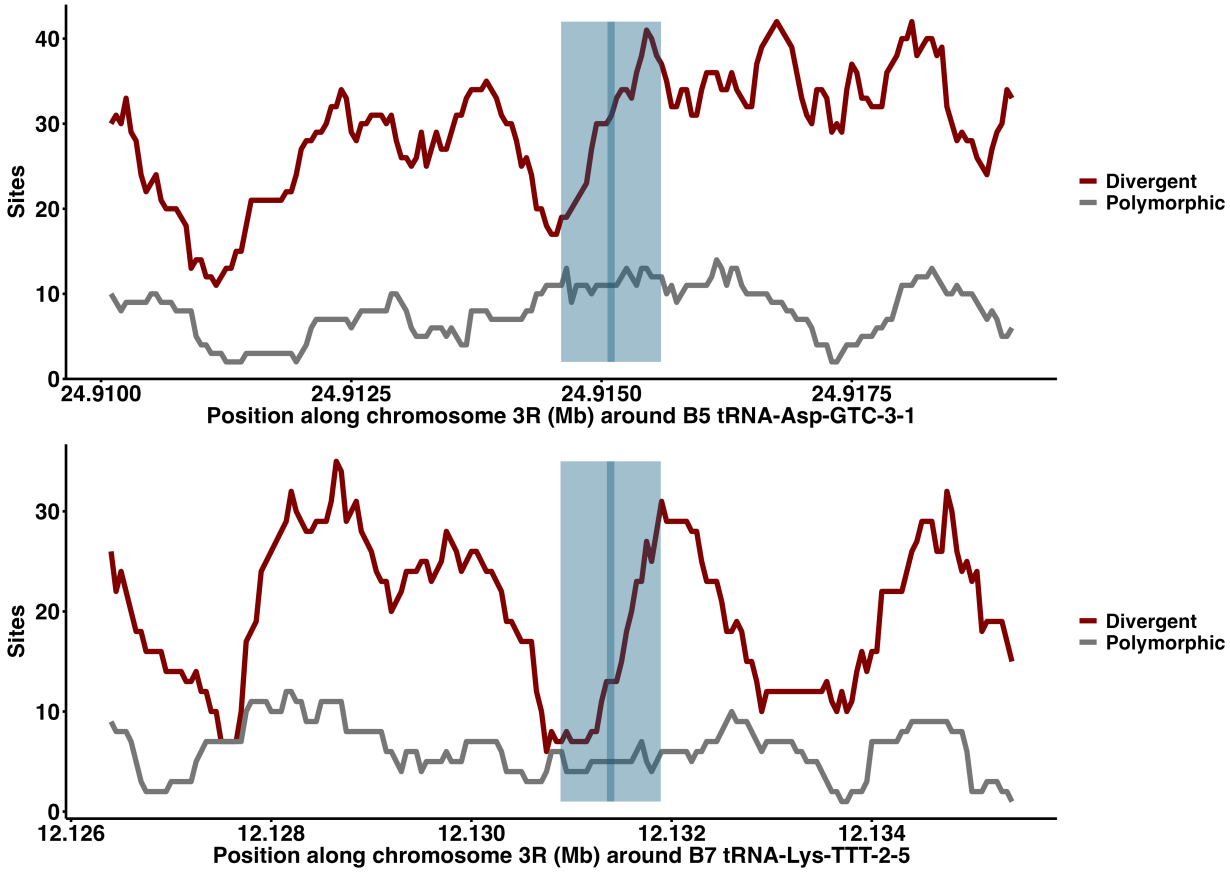



Figure 3.6: Two recently duplicated tRNAs (B7 and B5) *tRNA-Lys-TTT-2-5* and *tRNA-Asp-GTC-3-1* are in regions with significant reductions in proportions of segregating to divergent sites polarized with *D. simulans* and *D. yakuba* as compared to neutral expectations (shown in dark blue is the tRNA gene locus, light blue is the 1000bp window centered on the gene).

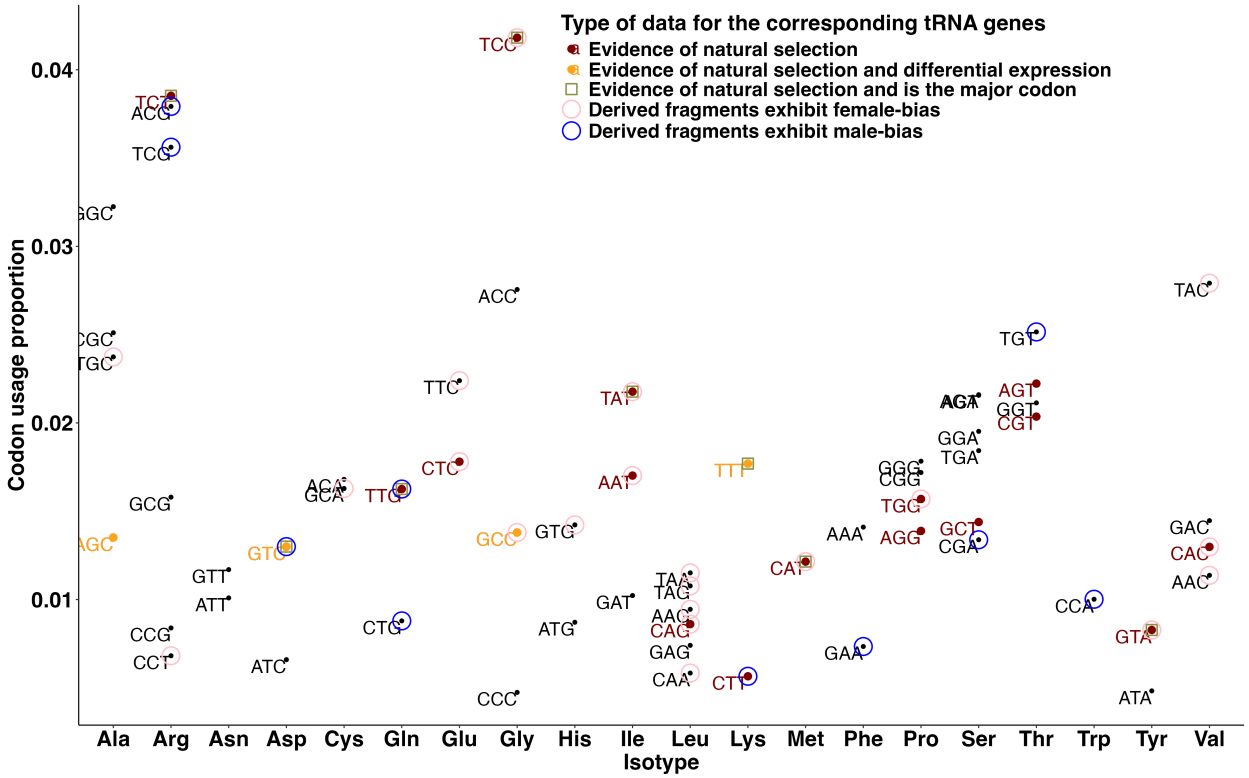



Figure 3.7: Anticodons of each transfer RNA isotype encoded by the nuclear *D. melanogaster* genome are plotted according to their isotype and frequency of usage in 12, 942 protein coding genes. Anticodons colored in red indicate significant Hudson-Kreitman-Aguadé-like test values, while those colored in yellow are differentially expressed in addition to having significant HKA-like results. Those anticodons marked by a green box have signatures of natural selection, as determined by statistical tests of neutrality, and are also the major codon of a given isotype. Blue and pink circles indicate sex-biased patterns of tRNA fragmentation in *D. melanogaster* testes or ovaries, respectfully.

## 3.4 Materials & Methods

### *3.4.1 Statistical tests of neutrality*

To test for natural selection at tRNA loci we first generated 10,000 neutral simulations following Hudson's method [Hudson, 2002] with 40,000 sites and 727 samples per simulation. For these simulations we used the following parameters: $N_0 = 1{,}000{,}000$ [Camacho et al., 2009], $\mu = 3.32 \times 10^{-9}$ [Wang et al., 2023b], and $\theta = 4 \cdot 1 \times 10^{6 \cdot 3.32 \times 10^{-9}}$ where $\theta = 4N_0\mu$.

We obtained empirical p-values for the Tajma's D and Fay-Wu's H statistics by identifying the number of simulated test values that were at least as extreme as observed values of D and H for all tRNA genes. We next polarized the *D. melanogaster* site frequency spectrum with population genetic data of D. simulans and D. yakuba to compute a Hudson-Kreitman-Aguadé-like test statistic. We used eight window and step configurations (400bp, 600bp, 800bp, 1kb and 25, 50bp) to collect observed polymorphic and divergent sites for comparison to neutral expectations [Sosa et al., 2025]. Only those tRNA loci that had significant HKA-like results ($p < 0.05$) in all window and step combinations after Benjamini & Hochberg correction were considered significant for downstream analysis [Benjamini and Hochberg, 1995].

### *3.4.2 tRNA extraction, library preparation, sequencing, and data analyses*

Total RNA was extracted in biological triplicate from adult Iso-1 *D. melanogaster* adult reproductive tissues (ovary, testis) using TRIzol (ThermoFisher cat. no. 15596026) following manufacturer instructions. Sequencing libraries were prepared using the multiplexed small RNA (MSR) protocol of Katanski et. al. and sequenced on Illumina platform [Watkins et al., 2022]. Briefly, Total RNA was deacylated using a one-pot periodate + deacylation reaction followed by beta-elimination and 3' end repair. Deacylated transfer RNAs were ligated to biotinylated oligos with barcodes and Illumina sequencing primers and incubated overnight at 16°C in 50 $\mu$L ligation solution. Ligation reaction solution was the following: 5% PEG 8000, 1$\mu$ NEB T4 RNA ligase I buffer, 50 $\mu$M ATP, 5% DMSO, 1 mM hexaamine cobalt (III) chloride, 0.8 $\mu$M barcode ligation oligo, and 1 U/$\mu$L NEB T4 RNA ligase I. The reaction was quenched with 50 $\mu$L of 100 mM EDTA and then combined with 8 $\mu$L of streptavidin coated MyOne C1 Dynabeads per sample. Samples were incubated with beads for 15 minutes at room temperature to ensure maximum binding. Samples were wash with 500 $\mu$L of high salt Tween (1 M NaCl, 0.1% Tween 20 and 20 mM Tris–HCl, pH 7.4) wash

buffer and once with 500 $\mu$L low salt (100 mM NaCl, 20 mM Tris–HCl, pH 7.4) wash buffer. Following ligation to the biotinylated beads and washing, multiple samples were combined at this point into a single tube. Samples were demethylated with AlkB using three stock solution prepared just prior to reaction: L-ascorbic acid 200 mM, 2-ketoglutarate 3 mM, and ammonium iron sulfate 5 mM. The final reaction mixture contained 2 mM L-ascorbic acid, 1 mM 2-ketoglutarate, 0.3 mM ammonium iron sulfate, 100 mM KCl, 50 mM MES pH 6, 50 ng/$\mu$L BSA, 4 $\mu$M wild-type AlkB, and 4 $\mu$M AlkB-D135S. 50 $\mu$L of the demethylase reaction mix was added to the samples and incubated at 37°C for 30 minutes. Samples were washed with 100 $\mu$L of high salt Tween (1 M NaCl, 0.1% Tween 20 and 20 mM Tris–HCl, pH 7.4) and 100 $\mu$L low salt (100 mM NaCl, 20 mM Tris–HCl, pH 7.4) wash buffer. The washed samples were then used for reverse transcription. 5 $\mu$L of Superscript IV master mix was added to each sample and incubated at 5°C for 10 minutes. The reaction was then incubated overnight at 35°C. 50 $\mu$L of RNase H reaction mix was added to each sample and incubated at 37°C for 15 minutes. Samples were washed and then resuspended in 40 $\mu$L of ddH2O. Sodium periodate mix was prepared by mixing 27 mg of NalO4 in 417 $\mu$L of ddH2O and 83 $\mu$L of 90 mM NaOAc/HOAc, pH 4.8. 10 $\mu$L of the periodate mix was added to the samples and then samples were incubated at room temperature for 30 minutes. Excess periodate was quenched with 10 $\mu$L of 1 mM ribose for 5 minutes. 50 $\mu$L of high salt Tween (1 M NaCl, 0.1% Tween 20 and 20 mM Tris–HCl, pH 7.4) and 100 $\mu$L low salt (100 mM NaCl, 20 mM Tris–HCl, pH 7.4) wash buffer was then added to each sample then washed with 100 $\mu$L of high salt Tween wash buffer and again with 100 $\mu$L of low salt wash buffer. A second ligation mix was prepared with 2 U/$\mu$l T4 RNA ligase I (NEB), 1x NEB T4 RNA ligase I buffer, 2 $\mu$M second ligation oligo, 25% PEG 8000, 50 $\mu$M ATP, 7.5% DMSO and 1 mM hexaamine cobalt chloride. 50 $\mu$L of this mix was added to each sample and then incubated overnight. Samples were diluted with 100 $\mu$L high salt Tween wash buffer then washed once with 100 $\mu$L high salt Tween wash buffer and once with 100 $\mu$L of low salt wash buffer. Beads

were resuspended in 7 $\mu$L of ddH2O per sample in the multiplexed tube. Following these treatments, beads were used as templates for PCR reaction and library preparation. Read alignment, quality control, abundance estimation, and statistical analyses were conducted using the tRAX software package [Holmes AD and TM, 2020].

### *3.4.3 Tissue specificity analysis*

The tissue specificity metric spm [Camargo, 2020, Pan et al., 2013] was computed using the log-transformed normalized counts matrix computed from the small RNA sequencing analysis for all tRNA reads.

$$X = (x_1,\ x_2, \ldots, x_i, \ldots, x_{n-1}, x_n)$$

$$SPM_i = \frac{x_i^2}{|x_i| * |X|}$$

Where X is a vector of expression values of a tissue 1 through n, the number of samples. SPM values range from 0 to 1, where larger values indicate more specific expression, and values $\geq 0.9$ are considered the most biased.

# CHAPTER IV

# TRANSLATIONAL FIDELITY OF GONADAL PROTEINS

## 4.1 Abstract

Messenger RNA translation is a fundamental process in all cellular life, yet it is not perfect. However, studies of translational error are limited both in number and scope due to technological limitations and challenges in matching small RNA expression data with proteomics. In this chapter, we test translational efficiency predictions by computationally determining the rate of translation errors in *D. melanogaster* gonads using publicly available mass spectrometry datasets from ovaries (n=4) and testes (n=2) and integrating that with our previous results from studying the evolution and expression of gonadal tRNAs. Approximately 4% of detected proteins in ovaries (3,696 total) and 2/% in testes (3,121 total) exhibited detectable erroneous peptides. Codon-specific error detection rates varied substantially across datasets and synonymous codons, revealing tissue-specific patterns of misincorporated amino acids. Contrary to expectations from translational selection theory, which predicts reduced error rates at highly expressed genes or those with optimized codons, no significant negative correlations were observed between mistranslation rates and these factors in either tissue. Further, both preferred and non-preferred codons showed similar error distributions, and codons lacking cognate tRNAs did not exhibit elevated errors, suggesting that ribosomal fidelity, tRNA modifications, or codon context may dominate over tRNA availability in modulating mistranslation. Most mistranslated proteins derived from ancient genes, with very few events in young genes, likely due to lower expression reducing detectability rather than inherent resistance to error. These findings challenge the role of translational selection in minimizing mistranslation in *Drosophila* gonads and highlight potential tissue-specific mechanisms influencing translational fidelity.

## 4.2 Introduction

Mistranslation refers to the erroneous incorporation of amino acids into polypeptides during protein synthesis. This phenomenon can occur through various mechanisms, including misacylation, where aminoacyl-tRNA synthetases attach the wrong amino acid to a tRNA, and misreading, where an incorrect tRNA pairs with a codon at the ribosome. These errors are not always random as tRNA modifications influence codon usage and genome structure, potentially exacerbating or mitigating mistranslation under specific conditions [Novoa et al., 2012]. Further, translational selection predicts that highly expressed and codon-optimized proteins should experience reduced translation error. Ultimately, while some mistranslation provides adaptive proteome diversity, excessive errors linked to tRNA dysfunction can lead to neurodegeneration or other disorders [Ishimura R, 2014, Netzer et al., 2009].

Studies of amino acid misincorporation (mistranslation) using mass spectrometry have predominantly been conducted in unicellular model organisms, particularly *E. coli* and *S. cerevisiae.* These systems allow high proteome coverage and sensitive detection of low-frequency substitution events due to their relatively simple proteomes, high protein abundance, and genetic tractability [Netzer et al., 2009]. Such studies have quantified error rates typically in the range of $10^{-4}$ to $10^{-2}$ per codon and linked them to codon usage, amino acid availability, and ribosomal accuracy [Hanson and Coller, 2018, Stikeleather et al., 2025, Landerer et al., 2024a]. In contrast, comprehensive, unbiased proteome-scale mass spectrometry analyses of amino acid misincorporation remain largely absent in multicellular eukaryotes, including metazoans such as *D. melanogaster* or mammals. While targeted studies have demonstrated specific mistranslation events—such as serine-for-proline substitutions from engineered tRNA variants in *Drosophila* at frequencies around 0.6% per codon—these rely on induced, non-physiological conditions rather than genome-wide profiling of natural mistranslation [Ishimura R, 2014]. The technical limitations in higher eukaryotes include greater

proteome complexity, tissue- and cell-type-specific tRNA expression and codon usage patterns [Dittmar et al., 2006, Pinkard et al., 2020], lower overall protein abundance in many tissues, developmental regulation of translation [Hudson et al., 2021], and reduced sequence coverage in mass spectrometry experiments, all of which hinder reliable detection of rare, stochastic substitution events across thousands of proteins. Consequently, mechanistic and quantitative insights into mistranslation in multicellular contexts have lagged behind those obtained from yeast and bacteria, despite growing evidence of its potential relevance in development, stress responses, and disease [Orellana et al., 2022, Zaborske et al., 2014].

In this chapter, we address these limitations by integrating genomic, transcriptomic, and proteomic data to comphrehensively characterize tRNA-mediated mistranslation in gonad tissues of *D. melanogaster*. Through computational analysis of publically available mass spectrometry datasets, we quantify the frequency and distribution of amino acid substitutions across the gonadal proteomes, linking them to tRNA expression profiles and codon usage patterns.

## 4.3 Results & Discussion

### *4.3.1 Landscape of Amino Acid Misincorporations at High Throughput*

We first accessed publicly available mass-spectrum datasets (testis: PXD041466, PXD044367; ovary: PXD012717,PXD036182,PXD037421,PXD043236) from experiments analyzing wild-type *D. melanogaster* ovaries and testes [ElMaghraby et al., 2019, Orellana et al., 2022]. This resulted in n=4 for ovaries, and n=2 for testes, underscoring the limited availability of sex-specific, multicellular proteomic data. Following the pipeline of Lander et. al. [Landerer et al., 2024b], we identified, processed, and filtered erroneous and reference peptides in each tissue relying on UniProt *D. melanogaster* protein annotations [Kong et al., 2017, Yu et al., 2020]. To detect errors in peptides, the method considers those peptides with a significant

mass shift compared to the reference protein sequence. In this analysis, we only consider those erroneous proteins which also have correct, detectable wild-type proteins. As post-translational modification can also affect peptide mass, quality filters were used to ensure modifications and other artifacts were not misidentified as translation errors. In total, 3,696 proteins were detected in ovary with about ~4 % having detectable error, while 3,121 were detected in testis and had about ~2% had detectable errors (Figs. 4.1, 4.2).

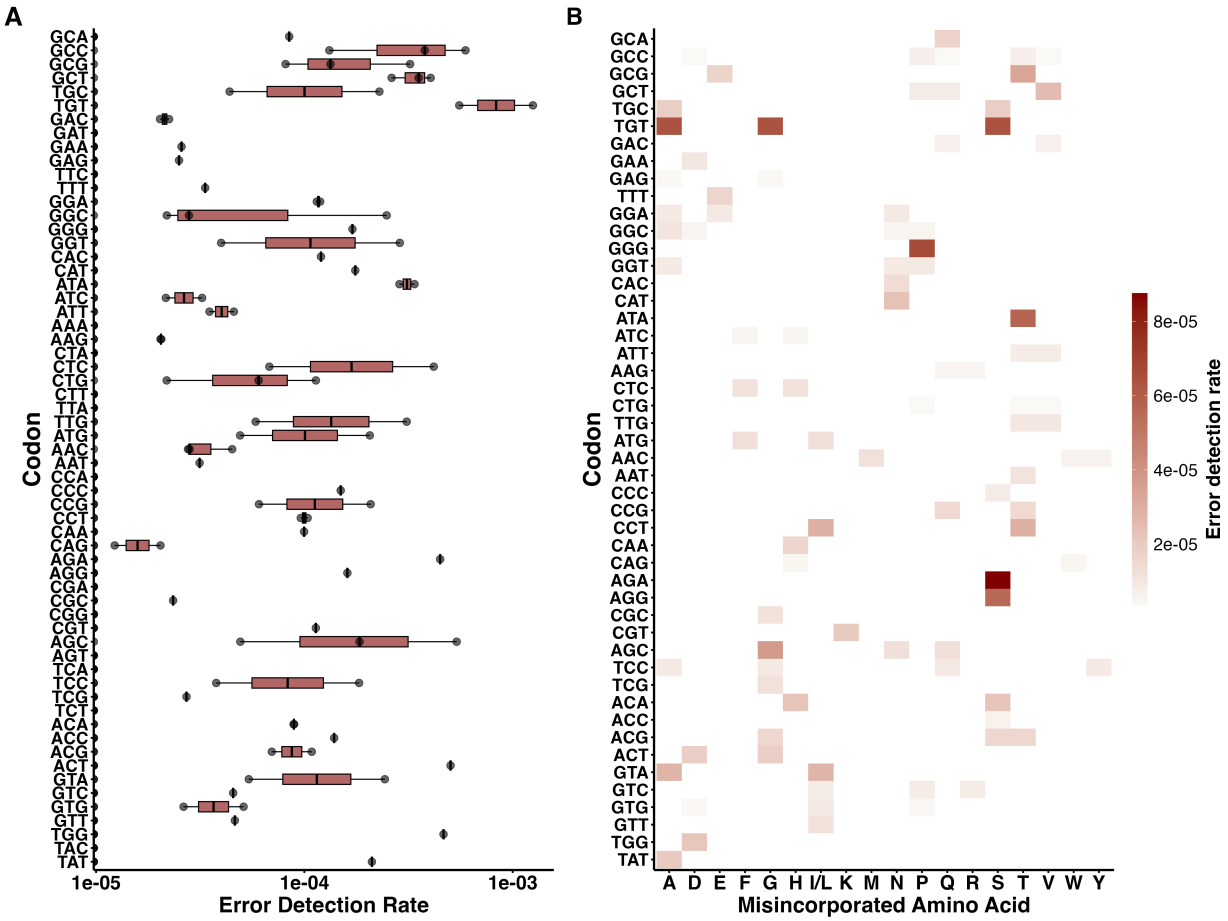



Figure 4.1: (A) Distribution of error detection rates across ovary proteomic datasets (ovary n = 4) for each codon observed sorted by isoaccepting group. Highly variable codon-specific error detection rates are observed across datasets and between synonymous codons. (B) Heatmap of ovary codon-specific error detection rates per misincorporated amino acid.

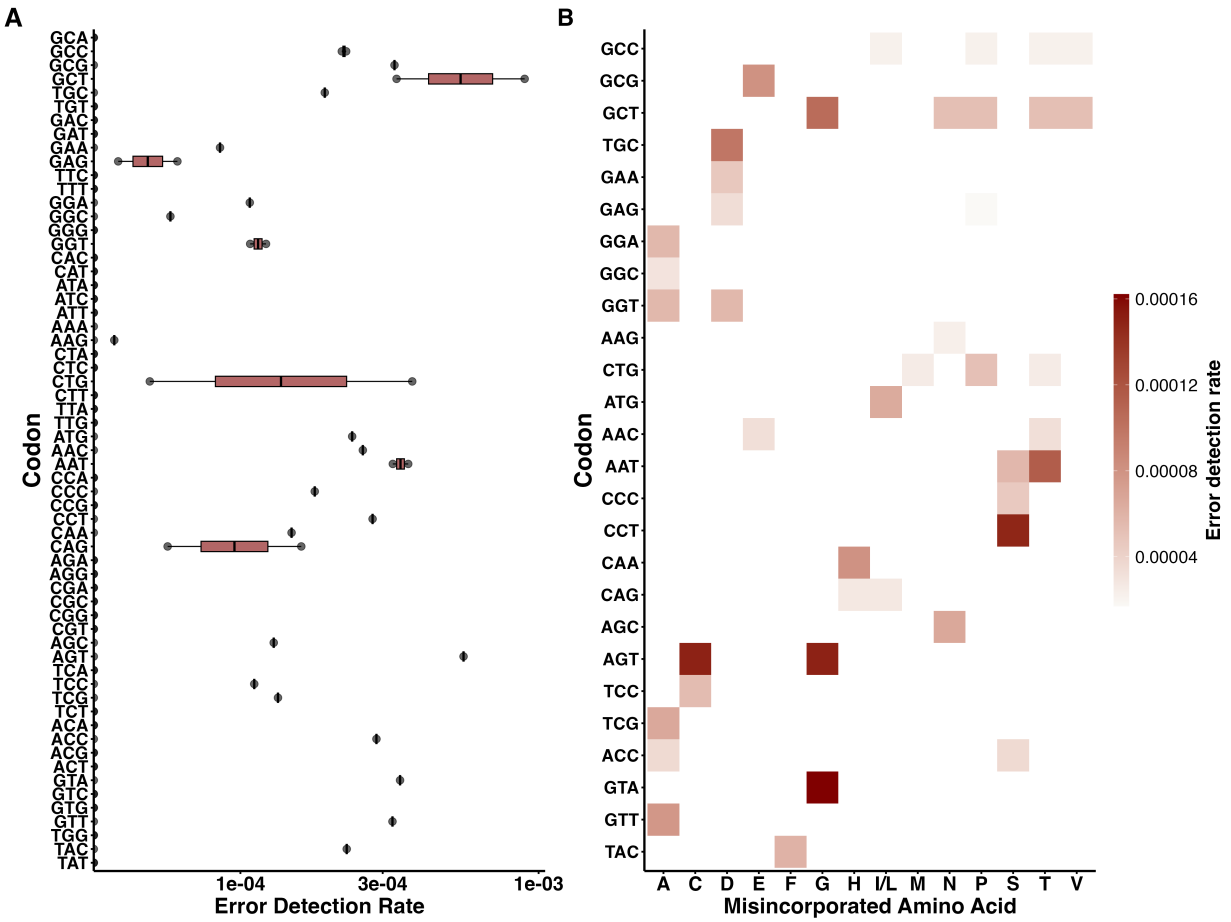



Figure 4.2: (A) Distribution of error detection rates across testis proteomic datasets (testis n = 2) for each codon observed sorted by isoaccepting group. Highly variable codon-specific error detection rates are observed across datasets and between synonymous codons. (B) Heatmap of testis codon-specific error detection rates per misincorporated amino acid.

As highly expressed genes are expected to have reduced translation error, we accessed RNA sequencing data of *D. melanogaster* ovaries and testes (PRJNA324423). Using these data, we correlated them with our detected error rates and previously computed $tAI_g$ values from Chapter II. In ovary, highly expressed genes are more biased in their synonymous codon usage ($R^2 = 0.308$, $p = 0.006$), but this was not observed in testis data ($R^2 = 0.045$, $p < 0.469$) (Fig. 4.3).

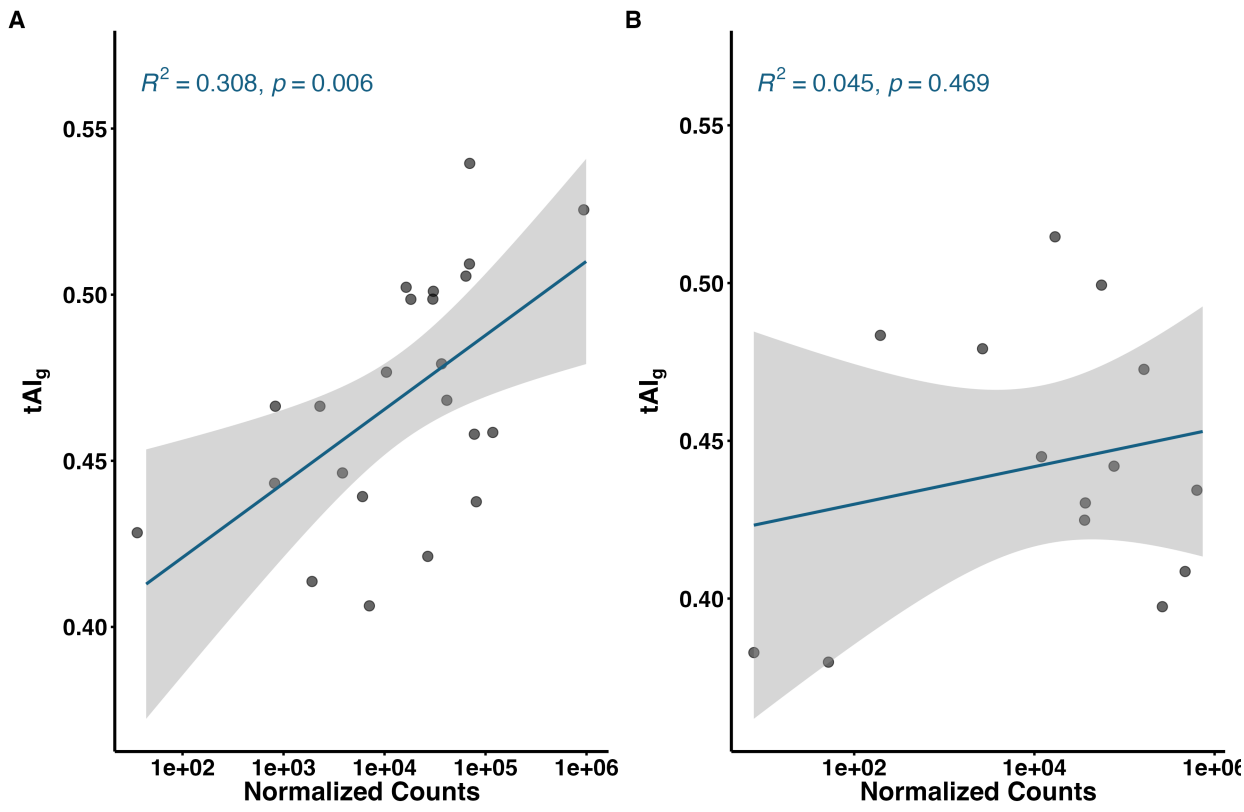



Figure 4.3: Correlation of mRNA expression of proteins with tAI in ovary (A) and testis (B). High expression is correlated with increased adaptation to the genomic tRNA pool in the ovary, but not testis.

### *4.3.2 Gonadal Mistranslation Rates Are Not Modulated by Translational Selection*

Under selection for translational efficiency, codon and tRNA copy numbers should be optimized such that translational error is reduced at highly expressed genes and preferred codons [Akashi, 1994, Bulmer, 1987, Ikemura, 1985, dos Reis et al., 2004, Bénitière et al., 2025, Yona et al., 2013]. We took our mistranslation observations and computed mean amino acid misincorporation per codon rates for ovary: $1.601171 \times 10^{-5}$ and testis: $5.964213 \times 10^{-5}$. These rates were computed using total errors at a given codon divided by the total number of the same codon observed in peptides derived from either gonad tissue. Despite the average error

rate being similar, the distribution of error rates across codons was significantly higher in testis compared to ovary ($p = 2.638 \times 10^{-15}$, rank-sum test). To directly test the relationship between mistranslation and translational selection, we correlated mean error rates per codon with expression, tRNA copy numbers and tAI values. Unexpectedly, we found that increased expression level is not correlated with reduced mistranslation in ovary (Fig. 4.4) ($R^2 = 0.126$, $p = 0.097$) or testis ($R^2 = 0.21$, $p < 0.0617$).

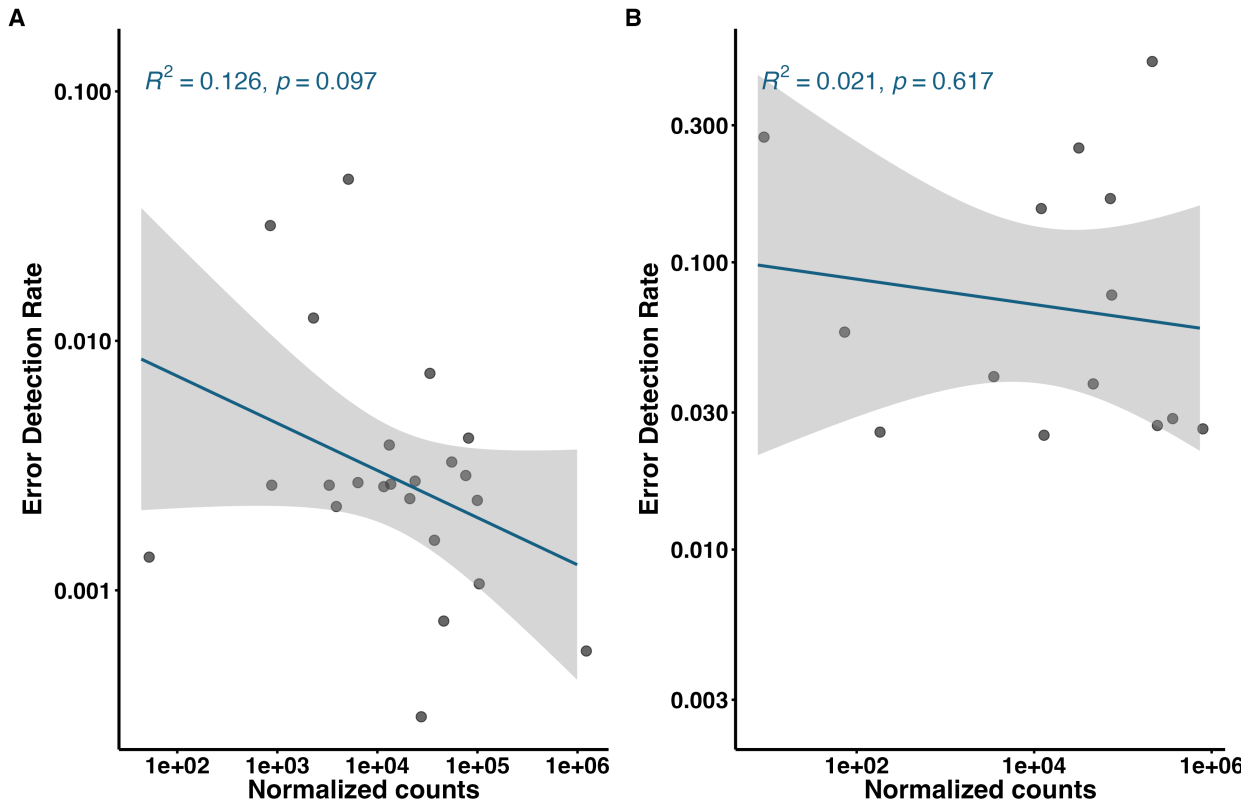



Figure 4.4: Correlation of mRNA expression of proteins with detected mistranslation error rate per codon in ovary (A) and testis (B).

A key aspect of translational efficiency is tRNA abundance and copy number [dos Reis et al., 2004, Hanson and Coller, 2018]. More available tRNAs should increase the opportunity for a tRNA to successfully be acylated by its cognate aaRS and codon [Landerer et al., 2024b]. As some codons lack cognate tRNAs, Table 1.1, [Rak et al., 2018], we wanted to determine if codons without corresponding tRNAs experienced increased translation error.

This hypothesis is in alignment with the translational selection model as non-optimal codons are expected to be translated more slowly and with increased error. However, in both gonad tissues we cannot identify significant differences in the distribution of mean error rate per codon and tRNA copy numbers per codon ($p = 0.5231$, and $p = 0.7347$, rank-sum test, Tables 4.1 and 4.2). These results suggest that the presence of a cognate tRNA does not necessarily reduce mistranslation rates, and that other factors such as codon context, tRNA modifications, or ribosomal fidelity may play more significant roles in determining error rates at specific codons. Further, in (Figs. 4.5 and 4.5) we show that increased tRNA copy number does not reduce mistranslation rates in either ovary or testis and that adaptation to the genomic tRNA pool does not necessarily reduce mistranslation rates. These data further suggest that tRNA abundance alone is not a major determinant of translational fidelity in gonadal tissues.

Table 4.1: Mean error rate per codon for peptides expressed in *D. melanogaster* ovaries.

| **Anticodon** | **Codon** | **AA** | **tRNA Copy Number** | **Mean Error Rate Per Codon** |
|---|---|---|---|---|
| TTT | AAA | Lys | 6 | 0 |
| GTT | AAC | Asn | 10 | 8.152400983994799e-6 |
| CTT | AAG | Lys | 13 | 4.974530404329831e-6 |
| ATT | AAT | Asn | 0 | 1.120410518413947e-5 |
| TGT | ACA | Thr | 6 | 2.221086999977789e-5 |
| GGT | ACC | Thr | 0 | 6.205783790492739e-6 |
| CGT | ACG | Thr | 3 | 1.565141175734051e-5 |
| AGT | ACT | Thr | 8 | 1.8616427135304193e-5 |
| TCT | AGA | Arg | 3 | 8.746610688358261e-5 |
| GCT | AGC | Ser | 6 | 2.0688000132403202e-5 |
| CCT | AGG | Arg | 3 | 5.572582892170521e-5 |

Table 4.1 ... *continued from previous page*

| **Anticodon** | **Codon** | **AA** | **tRNA Copy Number** | **Mean Error Rate Per Codon** |
|---|---|---|---|---|
| ACT | AGT | Ser | 0 | 0 |
| TAT | ATA | Ile | 2 | 5.721969501902555e-5 |
| GAT | ATC | Ile | 0 | 5.042940639545732e-6 |
| CAT | ATG | Met | 12 | 1.2316636080353733e-5 |
| AAT | ATT | Ile | 9 | 8.165531657766237e-6 |
| TTG | CAA | Gln | 4 | 1.5991300732401573e-5 |
| GTG | CAC | His | 5 | 1.3834128795739089e-5 |
| CTG | CAG | Gln | 8 | 5.036768409388536e-6 |
| ATG | CAT | His | 0 | 2.349624060150376e-5 |
| TGG | CCA | Pro | 5 | 0 |
| GGG | CCC | Pro | 0 | 8.241305422778968e-6 |
| CGG | CCG | Pro | 5 | 1.49774589243189e-5 |
| AGG | CCT | Pro | 7 | 2.9372026082359162e-5 |
| TCG | CGA | Arg | 10 | 0 |
| GCG | CGC | Arg | 0 | 1.1509334069930715e-5 |
| CCG | CGG | Arg | 0 | 0 |
| ACG | CGT | Arg | 10 | 2.0345052083333332e-5 |
| TAG | CTA | Leu | 2 | 0 |
| GAG | CTC | Leu | 0 | 1.1665344594279316e-5 |
| CAG | CTG | Leu | 8 | 3.617199058081365e-6 |
| AAG | CTT | Leu | 4 | 0 |
| TTC | GAA | Glu | 6 | 1.0426658881428034e-5 |
| GTC | GAC | Asp | 14 | 6.982167544092388e-6 |
| CTC | GAG | Glu | 14 | 3.9596276366170525e-6 |

Table 4.1 ... *continued from previous page*

| **Anticodon** | **Codon** | **AA** | **tRNA Copy Number** | **Mean Error Rate Per Codon** |
|---|---|---|---|---|
| ATC | GAT | Asp | 0 | 0 |
| TGC | GCA | Ala | 2 | 1.732531748644294e-5 |
| GGC | GCC | Ala | 0 | 5.17217811503663e-6 |
| CGC | GCG | Ala | 3 | 2.5093263295247338e-5 |
| AGC | GCT | Ala | 12 | 1.4164506579413305e-5 |
| TCC | GGA | Gly | 6 | 9.355324582986407e-6 |
| GCC | GGC | Gly | 14 | 6.853220465361082e-6 |
| CCC | GGG | Gly | 0 | 6.686279753944904e-5 |
| ACC | GGT | Gly | 0 | 9.260288180168167e-6 |
| TAC | GTA | Val | 2 | 2.757251571633396e-5 |
| GAC | GTC | Val | 0 | 8.117871494094248e-6 |
| CAC | GTG | Val | 7 | 6.0691856821840575e-6 |
| AAC | GTT | Val | 6 | 1.1856489056460601e-5 |
| GTA | TAC | Tyr | 10 | 0 |
| ATA | TAT | Tyr | 0 | 2.0356648481394024e-5 |
| TGA | TCA | Ser | 2 | 0 |
| GGA | TCC | Ser | 0 | 9.088265231932529e-6 |
| CGA | TCG | Ser | 4 | 1.2078315799644898e-5 |
| AGA | TCT | Ser | 8 | 0 |
| GCA | TGC | Cys | 7 | 1.87125748502994e-5 |
| CCA | TGG | Trp | 8 | 2.1723546151673798e-5 |
| ACA | TGT | Cys | 0 | 6.396315722144045e-5 |
| TAA | TTA | Leu | 4 | 0 |
| GAA | TTC | Phe | 8 | 0 |

Table 4.1 *... continued from previous page*

| Anticodon | Codon | AA | tRNA Copy Number | Mean Error Rate Per Codon |
|---|---|---|---|---|
| CAA | TTG | Leu | 4 | 1.024380249948781e-5 |
| AAA | TTT | Phe | 0 | 1.6602194810153904e-5 |

Table 4.2: Testis codon mistranslation rates.

| Anticodon | Codon | AA | tRNA Copy Number | Mean Error Rate Per Codon |
|---|---|---|---|---|
| TTT | AAA | Lys | 6 | 0 |
| GTT | AAC | Asn | 10 | 3.3024008454146164e-5 |
| CTT | AAG | Lys | 13 | 2.3456558453743668e-5 |
| ATT | AAT | Asn | 0 | 8.591065292096219e-5 |
| TGT | ACA | Thr | 6 | 0 |
| GGT | ACC | Thr | 0 | 3.61545970570158e-5 |
| CGT | ACG | Thr | 3 | 0 |
| AGT | ACT | Thr | 8 | 0 |
| TCT | AGA | Arg | 3 | 0 |
| GCT | AGC | Ser | 6 | 6.700167504187605e-5 |
| CCT | AGG | Arg | 3 | 0 |
| ACT | AGT | Ser | 0 | 1.5006002400960383e-4 |
| TAT | ATA | Ile | 2 | 0 |
| GAT | ATC | Ile | 0 | 0 |
| CAT | ATG | Met | 12 | 6.524008350730688e-5 |
| AAT | ATT | Ile | 9 | 0 |
| TTG | CAA | Gln | 4 | 8.068420203324189e-5 |
| GTG | CAC | His | 5 | 0 |

Table 4.2 ... *continued from previous page*

| **Anticodon** | **Codon** | **AA** | **tRNA Copy Number** | **Mean Error Rate Per Codon** |
|---|---|---|---|---|
| CTG | CAG | Gln | 8 | 2.7555041194786585e-5 |
| ATG | CAT | His | 0 | 0 |
| TGG | CCA | Pro | 5 | 0 |
| GGG | CCC | Pro | 0 | 4.616805170821791e-5 |
| CGG | CCG | Pro | 5 | 0 |
| AGG | CCT | Pro | 7 | 1.4821402104639098e-4 |
| TCG | CGA | Arg | 10 | 0 |
| GCG | CGC | Arg | 0 | 0 |
| CCG | CGG | Arg | 0 | 0 |
| ACG | CGT | Arg | 10 | 0 |
| TAG | CTA | Leu | 2 | 0 |
| GAG | CTC | Leu | 0 | 0 |
| CAG | CTG | Leu | 8 | 3.444860698445507e-5 |
| AAG | CTT | Leu | 4 | 0 |
| TTC | GAA | Glu | 6 | 4.6888920148168986e-5 |
| GTC | GAC | Asp | 14 | 0 |
| CTC | GAG | Glu | 14 | 2.575991756826378e-5 |
| ATC | GAT | Asp | 0 | 0 |
| TGC | GCA | Ala | 2 | 0 |
| GGC | GCC | Ala | 0 | 2.2246446130230696e-5 |
| CGC | GCG | Ala | 3 | 8.084074373484237e-5 |
| AGC | GCT | Ala | 12 | 6.312799200378767e-5 |
| TCC | GGA | Gly | 6 | 5.721805801911083e-5 |
| GCC | GGC | Gly | 14 | 3.0257186081694404e-5 |

Table 4.2 *... continued from previous page*

| Anticodon | Codon | AA | tRNA Copy Number | Mean Error Rate Per Codon |
|---|---|---|---|---|
| CCC | GGG | Gly | 0 | 0 |
| ACC | GGT | Gly | 0 | 5.7208237986270026e-5 |
| TAC | GTA | Val | 2 | 1.6204829039053638e-4 |
| GAC | GTC | Val | 0 | 0 |
| CAC | GTG | Val | 7 | 0 |
| AAC | GTT | Val | 6 | 7.753741180119407e-5 |
| GTA | TAC | Tyr | 10 | 6.044852807834129e-5 |
| ATA | TAT | Tyr | 0 | 0 |
| TGA | TCA | Ser | 2 | 0 |
| GGA | TCC | Ser | 0 | 5.4338966472857684e-5 |
| CGA | TCG | Ser | 4 | 6.682705159048383e-5 |
| AGA | TCT | Ser | 8 | 0 |
| GCA | TGC | Cys | 7 | 9.848335631278314e-5 |
| CCA | TGG | Trp | 8 | 0 |
| ACA | TGT | Cys | 0 | 0 |
| TAA | TTA | Leu | 4 | 0 |
| GAA | TTC | Phe | 8 | 0 |
| CAA | TTG | Leu | 4 | 0 |
| AAA | TTT | Phe | 0 | 0 |

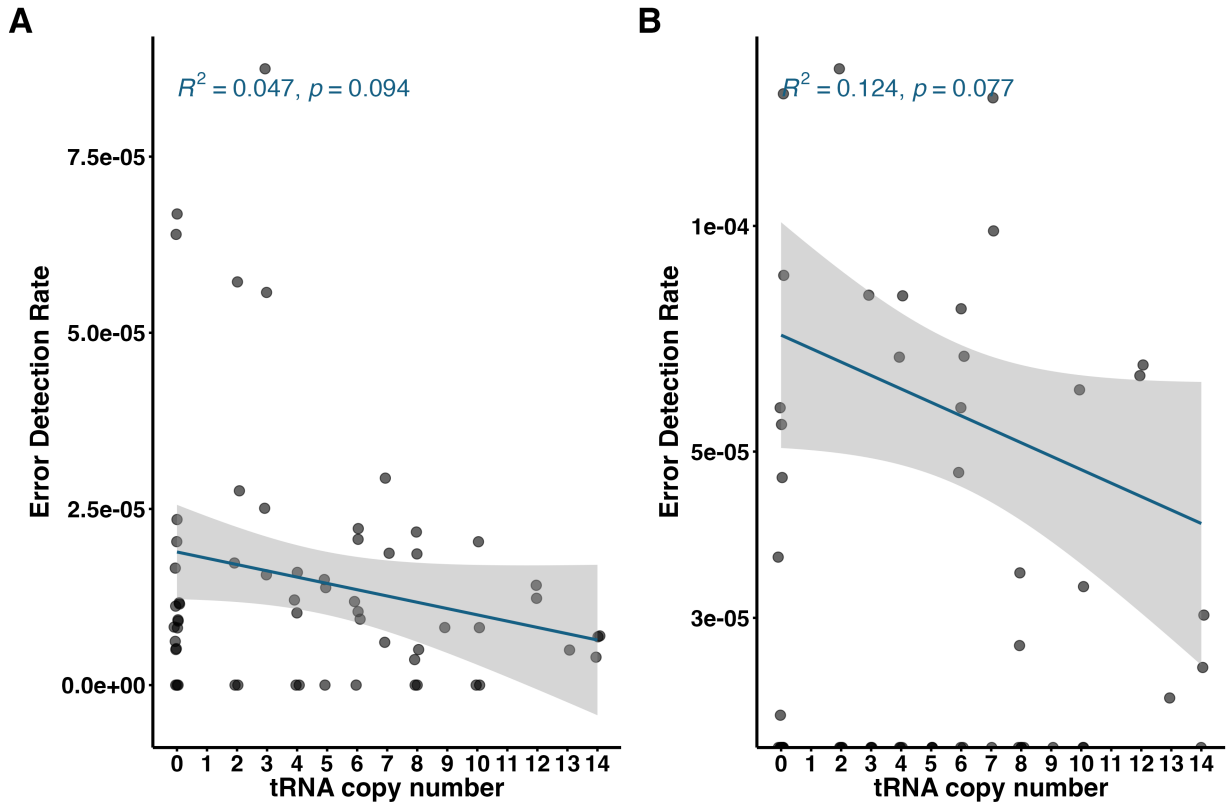



Figure 4.5: Increased tRNA copy number does not reduce misincorporation rates in ovary (A) and testis (B).

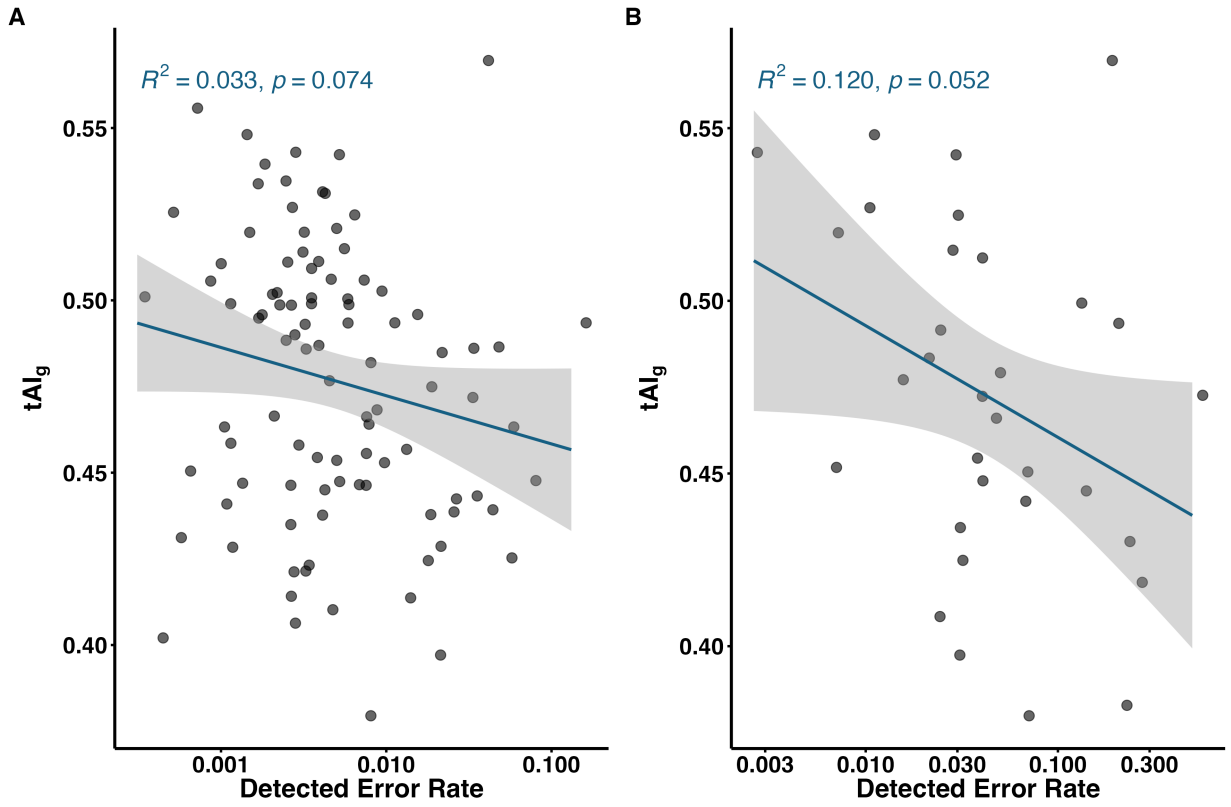



Figure 4.6: Correlation of tAI with detected mistranslation error rate per codon in ovary (A) and testis (B). No significant associaiton is observed in either tissue between biased synonymous codon usage adapted to the genomic tRNA pool and mistranslation error rates.

Using the gene age dating results from Chapter I, we investigated the relationship between coding gene age and expression. In (Fig. 4.7A) we show that in ovary expression rates are higher for proteins with detectable errors, indicating a relationship between expression and detectability. This is not observed in (Fig. 4.7C) where no significant difference in expression is observed between proteins with and without detected mistranslation events. In both tissues there no difference in the expression of proteins with errors between gene ages, (Fig. 4.7B,D). However, only one young gene was identified with a misincorporated amino acid: the B6 gene *FBgn0053910*.

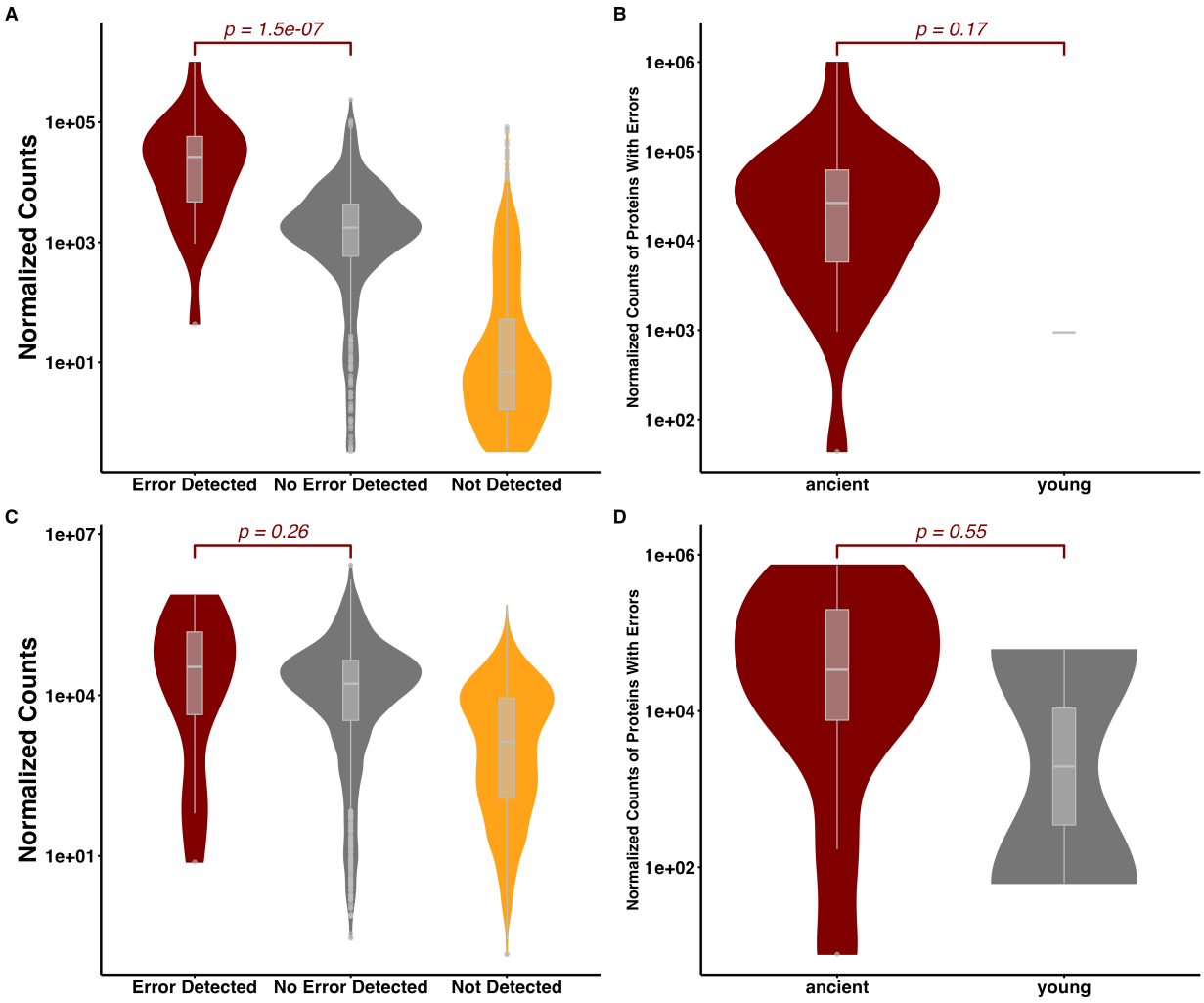



Figure 4.7: Comparison of expression distribution of genes expressed in *D. melanogaster* gonads grouped by mistranslation status. (A) Ancient genes have higher expression than young genes in testis both generally and when they have detected mistranslation events.

In Chapter II we defined preferred synomyous codons as the frequently used codon per amino acid. Assuming that preferred codons are more likely to be translated with higher fidelity, we statistically compared the distribution of error rates per codon between preferred and non-preferred codons. In both tissues, we found no significant difference in the distribution of error rates between preferred and non-preferred codons (ovary: $p = 0.2275$, testis: $p = 0.7512$, rank-sum test). This suggests that preferred codons are not necessarily translated with higher fidelity than non-preferred codons, and that other factors such as tRNA

modifications or ribosomal fidelity may play more significant roles in determining error rates at specific codons.

### *4.3.3 Potential Mechanistic Basis of Mistranslation in Gonads*

Amino acid misincorporation during translation arises primarily from nucleotide mismatches between codons in mRNA and anticodons in tRNA, leading to errors that compromise translational fidelity. The most common mechanism involves codon-anticodon mispairing, particularly at the wobble position (position 3 in codons), where non-standard base pairing allows near-cognate tRNAs to bind and insert incorrect amino acids. We categorized mistranslation events as combinations of mispaired nucleotides based on the position of the mispaired nucleotide in the codon (positions 1, 2, or 3). In both ovary and testis, we observed that the majority of mistranslation events involved mispairing at the wobble position (position 3). Overall, mispairing at positions 2 and 3 was the most common in both gonad tissues. Two mismatched nucleotides suggest a more complex mechanism of mistranslation, such as a combination of tRNA mischarging or misdecoding, consistent with previous reports of multiple error mechanisms contributing to mistranslation [Crick, 1966, Agris et al., 2018, Stikeleather et al., 2025]. The distribution of mispairing positions was similar between ovary and testis (Fig 4.8), suggesting that the underlying mechanisms of mistranslation may be conserved between gonadal tissues. However, ovary misincorporations do occur more frequently at all three codon positions (~30%) which is indicative of tRNA mischarging. Additionally, we found that certain amino acid substitutions were more prevalent than others, suggesting that specific codon-anticodon interactions may be more prone to misincorporation in gonadal tissues (Fig. 4.9).

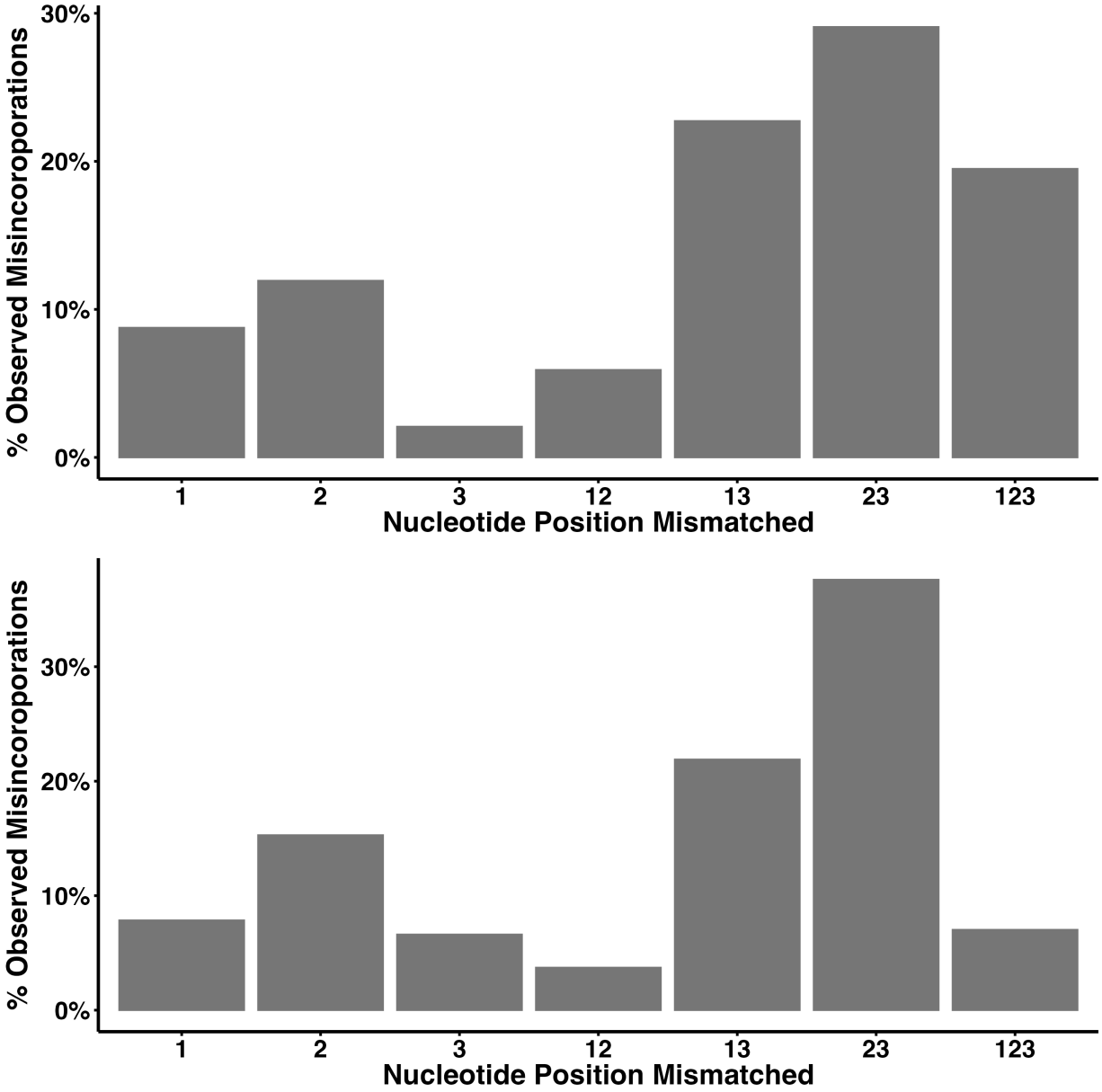



Figure 4.8: Types of mistranslation events based on the position of mispaired nucleotides in the codon between the original and misincorpoated residue (Top, ovary. Bottom, testis.).

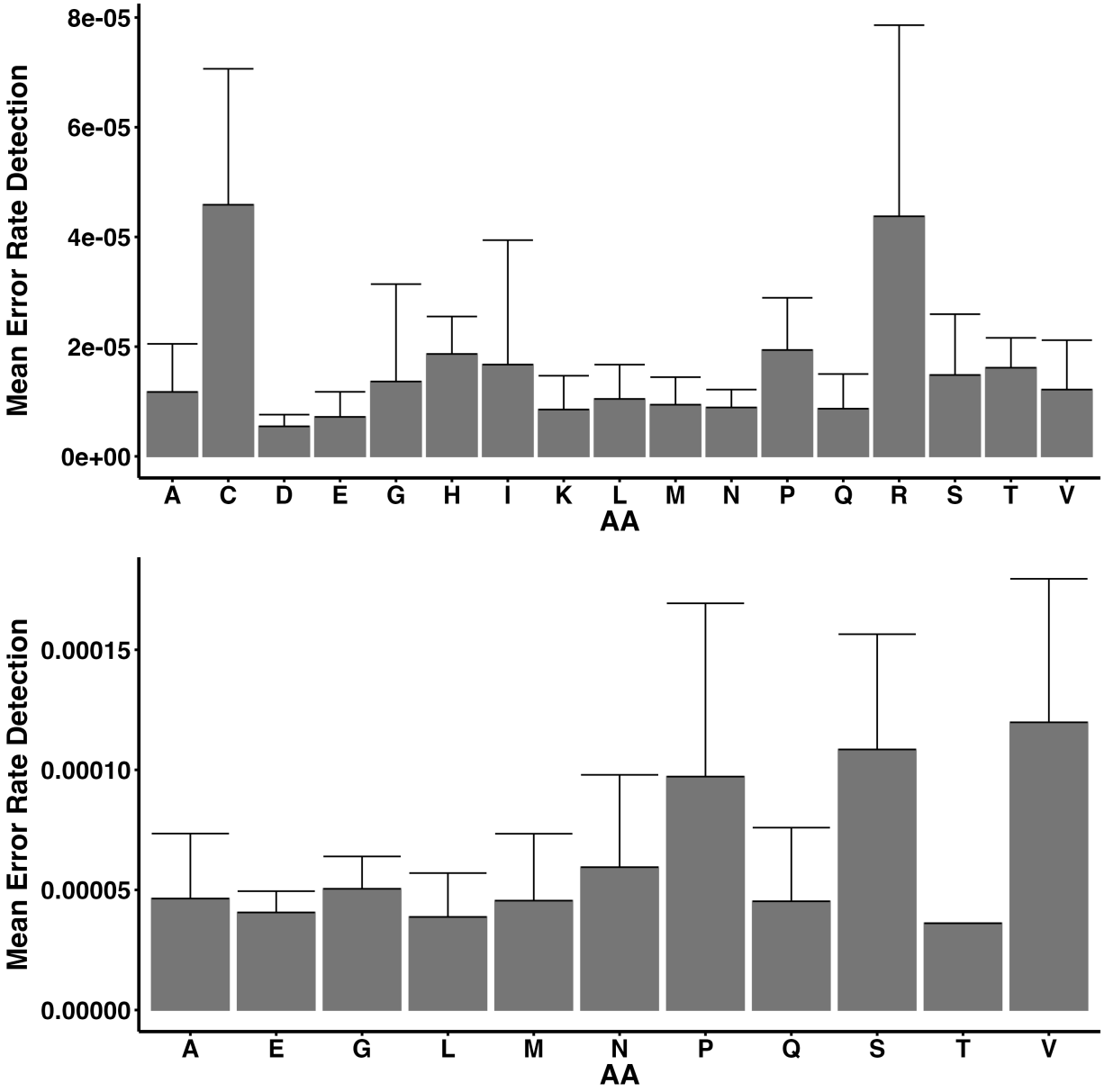



Figure 4.9: Tissue-specific error rate per amino acid (top, ovary; bottom, testis).

Finally, we intended to study gene age effects on mistranslation rates, but the limited number of young genes with detected mistranslation events precluded robust statistical analyses. In (Figs 4.10 and 4.11) we show the distribution of gene ages for proteins with (A) and without (B) detected mistranslation events in ovary (top) and testis (bottom), respectively. In both tissues, the majority of mistranslated proteins belong to ancient genes, with only one young gene identified in ovary and two in testis. This suggests two possibilities: that younger genes may be less prone to mistranslation, or lower expression levels in young genes generally may reduce the detectability of mistranslation events. Thus, given the limited number of young genes with detected errors we could not study young genes further. Future studies with larger datasets and improved detection methods may provide insights into how gene age influences mistranslation rates and mechanisms in gonadal tissues. Regardless, based on our findings from Chapter II, we had anticiapted increased mistranslation rates at codons of young genes. However, considering the results shown in (Figs. 4.4-4.6), unbiased codon usage may not be a significant driver of mistranslation rates in gonads.

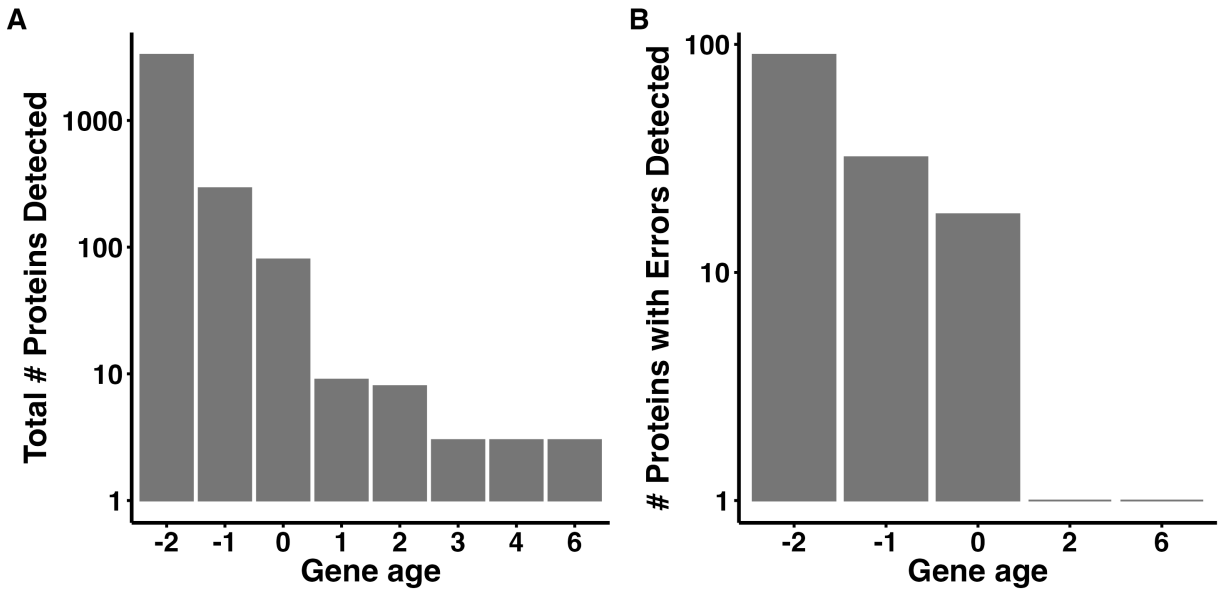



Figure 4.10: Age of proteins with detected (A) and those with mistranslation events (B) in ovary.

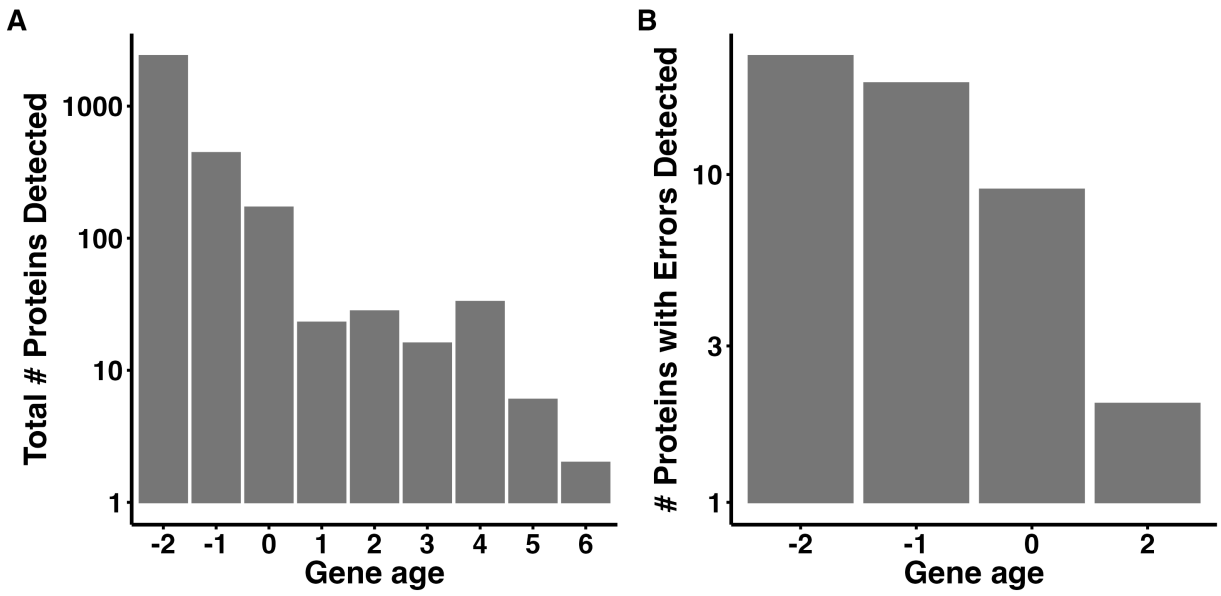



Figure 4.11: Age of proteins with detected (A) and those with mistranslation events(B) in testis.

## 4.4 Methods & Materials

### *4.4.1 Collection of proteomic data for D. melanogaster gonads.*

Mass spectrometry datasets for D. melanogaster gonads were identified and downloaded via ProteomeXchange for the following accessions: PXD041466, PXD044367, PXD012717, PXD036182, PXD037421, PXD043236. From each experiment only wild-type samples were retained for downstream analyses. In total, we analyzed 4 ovary and 2 testis proteomic datasets from multiple independent experiments.

### *4.4.2 Computing mistranslation rates from proteomic datasets*

Amino acid substitutions were identified following the method of Landerer [Landerer et al., 2024b]. Briefly, for each tissue's dataset, the corresponding raw mass spectrometry files was processed with MSFragger v.3.4 [Kong et al., 2017] in open search mode against the

*D. melanogaster* reference proteome UniProt Release 2025_04 to identify peptides with mass shifts corresponding to amino acid substitutions. The resulting peptide-spectrum matches (PSMs) were filtered to a 1% false discovery rate (FDR) using Philosopher v.4.0 [da Veiga Leprevost et al., 2020]. FDR-filtered PSMs were then aggregated and further filtered to retain only those peptides that could be unambiguously assigned to a single protein and had confident localization of any detected modifications. Among the remaining modified peptides, those with mass shifts corresponding to the exact difference between the original amino acid and a different amino acid were annotated as amino acid substitutions. For this purpose, leucine and isoleucine were treated as equivalent due to having the same mass. Error rates were calculated as the ratio of mistranslated peptides to the total number of peptides identified for each codon across all proteins.

### *4.4.3 Codon Usage and Translation Fidelity*

For all statistical comparisons of mistranslation rates, subsets were curated according to codon preferences and tRNA copy numbers identifed in Chapter II. Preferred codons were defined as the most frequently used codon for each amino acid in D. melanogaster, while non-preferred codons were defined as the less frequently used synonymous codons. We compared mistranslation rates between using the rank-sum test to determine if there was a significant difference in error rates based on multiple factors.

### *4.4.4 Characterization of translation error types in D. melanogaster*

Custom shell scripts were written to determine nucleotide positions of misincorporation events identified in the proteomic datasets. For each mistranslation event, we determined the codon position (1, 2, or 3) of the misincorporated amino acid and categorized the type of mispairing (e.g., A-G, C-U) based on the original codon and the amino acid substitution observed. We then analyzed the distribution of mispairing types and positions in both ovary

and testis datasets to identify potential mechanistic differences in mistranslation between gonadal tissues. Additionally, we compared error rates between preferred and non-preferred codons to assess whether codon usage bias influences translational fidelity in gonads.

# CHAPTER V

# CONCLUSION

Coordination of cytosolic tRNA availability and translational demand is essential to Darwinian fitness and cellular functionality across all domains of life [Hughes et al., 2023, Orellana et al., 2022]. Yet, in the preceding chapters I have demonstrated that the supply of translational machinery is neither conserved between species nor optimized in copy number to solely match global codon usage.

## 5.1 Transfer RNA Evolutionary Dynamics

Despite performing essentials role in mRNA translation and protein biosynthesis, scant attention has been paid to the evolutionary genetics of tRNA which have been revealed to vary significantly in copy number and isoacceptor composition between closely related taxa and across phylogenetic domains [Parisien et al., 2013, Rogers et al., 2010, Tang et al., 2009]. We utilized the genomes of 24 Drosophilid species spanning a 60Ma time range to test our hypotheses that tRNA gene repertoires can change rapidly, similar to the proteome, and to ask when tRNA genes emerge, how they expand and contract; and what evolutionary forces are maintaining seemingly redundant copies in eukaryotic genomes. Our investigations of the tRNA gene repertoire expansion in the family Drosophilidæ provide evidence of recently evolved tRNA genes as well as the continuous emergence of lineage-specific tRNA genes over evolutionary time. Through comparative genomic analyses we detected 26.46% of *D. melanogaster* tRNA genes to be *Sophopohpora* subgenus-specific, 1.03% to be *melanogaster* subgroup-specific, and one gene that is specific to *D. melanogaster*, suggesting an unexpectedly high rate of generation and retention of tRNA genes despite their deeply conserved translational function. Taxonomically restricted and isodecoding tRNA duplicates originating in recently diverged Drosophilid species exhibit patterns of sexually dimorphic functions

and characteristics such as sex-biased abundance, fragmentation, and modification. The preservation of seemingly redundant tRNA gene copies is likely influenced by population genetics, nearly neutral processes, and sexual selection-not solely environmental factors (e.g., nutrient-driven tRNA modification) or directional selection for translational efficiency and accuracy, particularly at the genome-wide level [Charlesworth et al., 1994, Zaborske et al., 2014]. Our work uncovers basic patterns of tRNA gene evolutionary biology and maintenance of new isoaccepting and isodecoding tRNA copies, suggesting there may be more factors contributing to the extreme copy number variation of tRNA genes between metazoan species than previously anticipated.

## 5.2 Translational Conflict Hypothesis

New gene emergence in *Drosophila* [Zhang et al., 2010a, Zhou et al., 2008] through various mechanisms (*e.g.* RNA- or DNA-based duplication, *de novo,* reading frame shifts) coupled with sex- or tissue-specific expression [Long et al., 2013, VanKuren and Long, 2018, Xia et al., 2021] may have caused the protein-coding genome to evolve in a manner such that the evolution of new tRNA genes was necessary to respond to novel translational demands in disparate tissue environments [Allen et al., 2022, Dittmar et al., 2006, Otto Hagenbuchle and Sprague, 1979, Pinkard et al., 2020]. Thus, we anticipated that tRNA genes may have been co-evolving with the emergence of new genes to translate young, codon-unoptimized genes to ensure translation of fthe ever-changing proteome. In stark contrast, our work revealed the existence of translational conflict between young genes (duplicates, orphans) and the genomic tRNA pool of *D. melanogaster*. In this dissertation I define translational conflict as the conflict between unbiased codon usage of recently evovled genes and the codon co-evolved tRNA repertoire (Fig. 5.1).

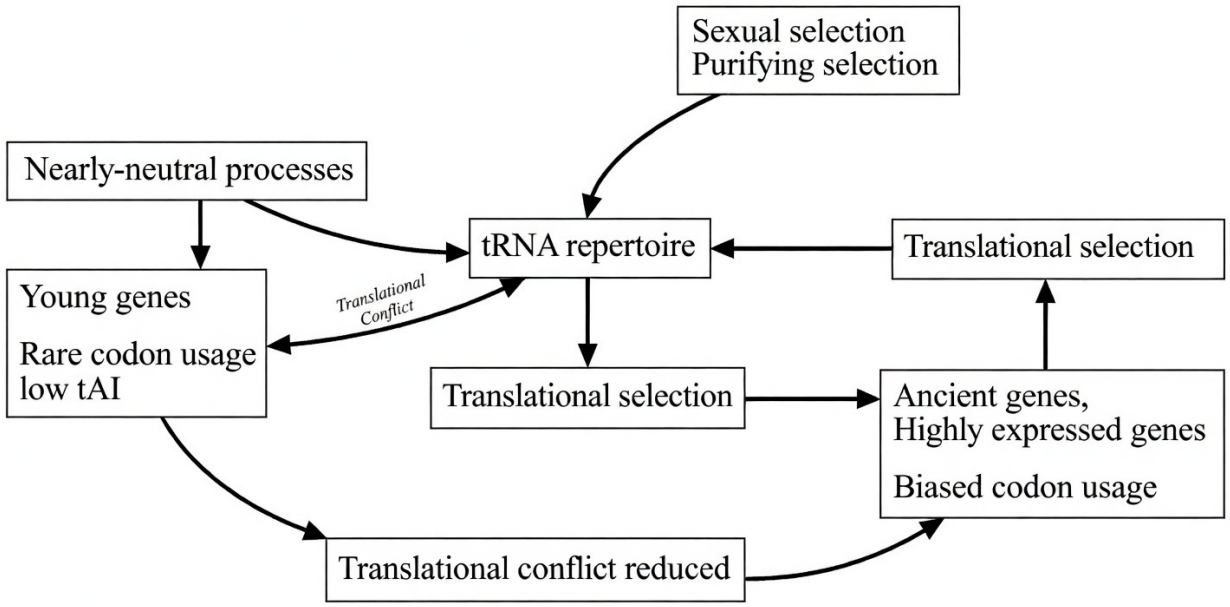



Figure 5.1: Schematic of the proposed model of metazoan tRNA evolution. The model posits that tRNA gene repertoires evolve under multiple selective pressures, both adaptive and not which shape their composition and abundance over time. While translational selection may weakly affect some coding genes which affects tRNA copy number or expression in certain contexts, nearly neutral processes and sexual selection may also play significant roles in shaping tRNA gene repertoires. The emergence of new genes with unbiased codon usage creates a conflict with the existing tRNA pool, which has co-evolved to match the codon usage of older genes. This conflict may drive the evolution of new tRNA gene duplicates that are better suited to translate the codons of young genes, particularly in tissues where these genes are highly expressed, while other likely ancient copies maintain translational functions. The resolution of this conflict through the evolution of new tRNA genes may contribute to the maintenance and functional integration of young genes into the genome, ultimately enhancing organismal fitness either through translation fidelity or non-canonical functions.

In this model of the translational conflict hypothesis, I assume that tRNA gene complements are influenced by several forces and contexts, not just seelction to match abundant codons. The emergence of new genes with unbiased codon usage creates a conflict with the existing tRNA pool, which has co-evolved to match the codon usage of older genes. This conflict may drive the evolution of new tRNA gene duplicates that are better suited to translate the codons of young genes, particularly in tissues where these genes are highly expressed, while other likely ancient copies maintain translational functions. The resolution

of this conflict through the evolution of new tRNA genes may contribute to the maintenance and functional integration of young genes into the genome, ultimately enhancing organismal fitness either through translation fidelity or non-canonical functions. Recent studies have begun to experimentally investigate the translational efficiencies of different tissues such as the work by Allen et. al. which showed that the brain and testes of *D. melanogaster* express a young gene, *(RpL10Aa)*, biased for rare codons [Allen et al., 2022] which causes disrupted female fertility when its codons are optimized experimentally. Their findings highlight a new and important mechanism for how translational conflict may regulate sex-specific expression. These results are also consistent with observations that new genes are typically not codon optimized and often have sex-tissue biased expression as reported in previous studies of plants and animals [Emerson et al., 2004, VanKuren and Long, 2018, Zhang et al., 2019, 2011, 2010a,b]. Thus, translational conflict may be a feature of new gene origination across species [Zhang et al., 2019].

# BIBLIOGRAPHY


P. F. Agris, E. R. Eruysal, A. Narendran, V. Y. P. Väre, S. Vangaveti, and S. V. Ranganathan. Celebrating wobble decoding: Half a century and still much is new. *RNA Biol*, 15(4-5):537–553, 2018. ISSN 1547-6286 (Print) 1547-6286. doi:10.1080/15476286.2017.1356562.

Hiroshi Akashi. Synonymous codon usage in drosophila melanogaster: Natural selection and translational accuracy. *Genetics*, 136:927–935, 1994.

Yasutoshi Akiyama and Pavel Ivanov. trna-derived rnas: Biogenesis and roles in translational control. *WIREs RNA*, 14(6):e1805, 2023. ISSN 1757-7004. doi:https://doi.org/10.1002/wrna.1805.

S. R. Allen, R. K. Stewart, M. Rogers, I. J. Ruiz, E. Cohen, A. Laederach, C. M. Counter, J. K. Sawyer, and D. T. Fox. Distinct responses to rare codons in select drosophila tissues. *Elife*, 11, 2022. ISSN 2050-084X (Electronic) 2050-084X (Linking). doi:10.7554/eLife.76893.

Sangpen An and Karin Musier-Forsyth. Trans-editing of cys-trnapro by haemophilus influenzae ybak protein. *J Biol Chem*, 279(41):42359–42362, 2004. doi:10.1074/jbc.C400304200.

G. B. Ayan, H. J. Park, and J. Gallie. The birth of a bacterial trna gene by large-scale, tandem duplication events. *Elife*, 9, 2020. ISSN 2050-084X (Electronic) 2050-084X (Linking). doi:10.7554/eLife.57947.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995. ISSN 0035-9246.

E. Betran, K. Thornton, and M. Long. Retroposed new genes out of the x in drosophila. *Genome Res*, 12(12):1854–9, 2002. ISSN 1088-9051 (Print) 1088-9051 (Linking). doi:10.1101/gr.6049.

Michael Bulmer. Coevolution of codon usage and transfer rna abundance. *Nature*, 325(19): 728–730, 1987.

F. Bénitière, T. Lefébure, and L. Duret. Variation in the fitness impact of translationally optimal codons among animals. *Genome Res*, 35(3):446–458, 2025. ISSN 1088-9051 (Print) 1088-9051. doi:10.1101/gr.279837.124.

G. Caetano-Anolles and F. J. Sun. The natural history of transfer rna and its interactions with the ribosome. *Front Genet*, 5:127, 2014. ISSN 1664-8021 (Print) 1664-8021 (Electronic) 1664-8021 (Linking). doi:10.3389/fgene.2014.00127.

Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. Blast+: architecture and applications. *BMC Bioinformatics*, 10(1):421, 2009. ISSN 1471-2105. doi:10.1186/1471-2105-10-421.

Vasconcelos A. P. Fiamenghi M. B. Camargo, A. P. tspex: a tissue-specificity calculator for gene expression data, preprint (version 1). *Research Square*, 2020. doi:https://doi.org/10.21203/rs.3.rs-51998/v1.

P. P. Chan, B. Y. Lin, A. J. Mak, and T. M. Lowe. trnascan-se 2.0: improved detection and functional classification of transfer rna genes. *Nucleic Acids Res*, 49(16): 9077–9096, 2021. ISSN 1362-4962 (Electronic) 0305-1048 (Print) 0305-1048 (Linking). doi:10.1093/nar/gkab688.

Patricia P. Chan and Todd M. Lowe. Gtrnadb 2.0: an expanded database of transfer rna genes identified in complete and draft genomes. *Nucleic Acids Research*, 44(D1):D184–D189, 2015. ISSN 0305-1048. doi:10.1093/nar/gkv1309.

Karen B. Chapman and et al. Initiation of retrovirus cdna synthesis: The role of trna primers. *J Biol Chem*, 267(13):8867–8874, 1992.

B. Charlesworth, P. Sniegowski, and W. Stephan. The evolutionary dynamics of repetitive dna in eukaryotes. *Nature*, 371(6494):215–20, 1994. ISSN 0028-0836 (Print) 0028-0836. doi:10.1038/371215a0.

Q. Chen, M. Yan, Z. Cao, X. Li, Y. Zhang, J. Shi, G. H. Feng, H. Peng, X. Zhang, Y. Zhang, J. Qian, E. Duan, Q. Zhai, and Q. Zhou. Sperm tsrnas contribute to intergenerational inheritance of an acquired metabolic disorder. *Science*, 351(6271):397–400, 2016. ISSN 0036-8075. doi:10.1126/science.aad7977.

S. Chen, B. H. Krinsky, and M. Long. New genes as drivers of phenotypic evolution. *Nat Rev Genet*, 14(9):645–60, 2013. ISSN 1471-0064 (Electronic) 1471-0056 (Linking). doi:10.1038/nrg3521.

Shifu Chen, Yanqing Zhou, Yaru Chen, and Jia Gu. fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics*, 34(17):i884–i890, 2018. ISSN 1367-4803.

Naoko Takezaki Claudia A. M. Russo and Masatoshi Nei. Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.*, 12(3):391–404, 1995.

Alexander L Cope and Premal Shah. Evolution of trna pool shapes variation in selection on codon usage across the saccharomycotina subphylu. *bioRxiv*, page 2024.09.27.615277, 2024. doi:10.1101/2024.09.27.615277.

J. M. Coughlan, A. J. Dagilis, A. Serrato-Capuchina, H. Elias, D. Peede, K. Isbell, D. M. Castillo, B. S. Cooper, and D. R. Matute. Patterns of population structure and introgression among recently differentiated drosophila melanogaster populations. *Mol Biol Evol*, 39(11), 2022. ISSN 1537-1719 (Electronic) 0737-4038 (Print) 0737-4038 (Linking). doi:10.1093/molbev/msac223.

F. H. Crick. On protein synthesis. *Symp Soc Exp Biol*, 12:138–63, 1958. ISSN 0081-1386 (Print) 0081-1386.

F. H. C. Crick. Codon-anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology*, 19:548–555, 1966.

Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970. ISSN 0028-0836.

Francis HC Crick. The origin of the genetic code. *Journal of molecular biology*, 38(3): 367–379, 1968. ISSN 0022-2836.

Felipe da Veiga Leprevost, Sarah E Haynes, Dmitry M Avtonomov, Hui-Yin Chang, Avinash K Shanmugam, Dattatreya Mellacheruvu, Andy T Kong, and Alexey I Nesvizhskii. Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nature methods*, 17(9):869–870, 2020. ISSN 1548-7091.

T. Dale and O. C. Uhlenbeck. Amino acid specificity in translation. *Trends Biochem Sci*, 30(12):659–65, 2005. ISSN 0968-0004 (Print) 0968-0004 (Linking). doi:10.1016/j.tibs.2005.10.006.

Charles Darwin. *Journal of researches into the natural history and geology of the countries visited during the voyage of HMS" Beagle" round the world: under the command of Capt. Fitz Roy, RN*, volume 1. Ward, Lock and Company, 1889.

G de Sena Brandine and AD Smith. Falco: high-speed fastqc emulation for quality control of sequencing data [version 2; peer review: 2 approved]. *F1000Research*, 8(1874), 2021. doi:10.12688/f1000research.21142.2.

K. A. Dittmar, J. M. Goodenbour, and T. Pan. Tissue-specific differences in human transfer rna expression. *PLoS Genet*, 2(12):e221, 2006. ISSN 1553-7404 (Electronic) 1553-7390 (Print) 1553-7390 (Linking). doi:10.1371/journal.pgen.0020221.

Chuan Dong, Shengqian Xia, Li Zhang, Deanna Arsala, Chengchi Fang, Shengjun Tan, Andrew G Clark, and Manyuan Long. Subcellular enrichment patterns of new genes in drosophila evolution. *Molecular Biology and Evolution*, 42(2), 2025. ISSN 1537-1719. doi:10.1093/molbev/msaf038.

M. dos Reis, R. Savva, and L. Wernisch. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res*, 32(17):5036–44, 2004. ISSN 1362-4962 (Electronic) 0305-1048 (Print) 0305-1048 (Linking). doi:10.1093/nar/gkh834.

Mostafa F. ElMaghraby, Peter Refsing Andersen, Florian Pühringer, Ulrich Hohmann, Katharina Meixner, Thomas Lendl, Laszlo Tirian, and Julius Brennecke. A heterochromatin-specific rna export pathway facilitates pirna production. *Cell*, 178(4): 964–979.e20, 2019. ISSN 0092-8674. doi:10.1016/j.cell.2019.07.007.

J. J. Emerson, H. Kaessmann, E. Betrán, and M. Long. Extensive gene traffic on the mammalian x chromosome. *Science*, 303(5657):537–40, 2004. ISSN 0036-8075. doi:10.1126/science.1090042.

M. J. Ford and C. F. Aquadro. Selection on x-linked genes during speciation in the drosophila athabasca complex. *Genetics*, 144(2):689–703, 1996. ISSN 0016-6731 (Print) 0016-6731. doi:10.1093/genetics/144.2.689.

K. Fujishima and A. Kanai. trna gene diversity in the three domains of life. *Front Genet*, 5:142, 2014. ISSN 1664-8021 (Print) 1664-8021 (Electronic) 1664-8021 (Linking). doi:10.3389/fgene.2014.00142.

R. Geslain and T. Pan. Functional analysis of human trna isodecoders. *J Mol Biol*, 396 (3):821–31, 2010. ISSN 1089-8638 (Electronic) 0022-2836 (Print) 0022-2836 (Linking). doi:10.1016/j.jmb.2009.12.018.

R. Giege. Toward a more complete view of trna biology. *Nat Struct Mol Biol*, 15(10):1007–14, 2008. ISSN 1545-9985 (Electronic) 1545-9985 (Linking). doi:10.1038/nsmb.1498.

Sophie Giguère, Xuesong Wang, Sabrina Huber, Liling Xu, John Warner, Stephanie R. Weldon, Jennifer Hu, Quynh Anh Phan, Katie Tumang, Thavaleak Prum, Duanduan Ma, Kathrin H. Kirsch, Usha Nair, Peter Dedon, and Facundo D. Batista. Antibody production relies on the trna inosine wobble modification to meet biased codon demand. *Science*, 383 (6679):205–211, 2024. doi:doi:10.1126/science.adi1763.

H. Goodarzi, H. C. B. Nguyen, S. Zhang, B. D. Dill, H. Molina, and S. F. Tavazoie. Modulated expression of specific trnas drives gene expression and cancer progression. *Cell*, 165 (6):1416–1427, 2016. ISSN 0092-8674 (Print) 0092-8674. doi:10.1016/j.cell.2016.05.046.

Hani Goodarzi, Xinran Liu, H. C. Brandon Nguyen, Steven Zhang, Lisa Fish, and Sohail F. Tavazoie. Endogenous trna-derived fragments suppress breast cancer progression via ybx1 displacement. *Cell*, 161(4):790–802, 2015. doi:10.1016/j.cell.2015.02.053.

J. M. Goodenbour and T. Pan. Diversity of trna genes in eukaryotes. *Nucleic Acids Res*, 34 (21):6137–46, 2006. ISSN 1362-4962 (Electronic) 0305-1048 (Print) 0305-1048 (Linking). doi:10.1093/nar/gkl725.

C. Guijarro-Clarke, P. W. H. Holland, and J. Paps. Widespread patterns of gene loss in the evolution of the animal kingdom. *Nat Ecol Evol*, 4(4):519–523, 2020. ISSN 2397-334X (Electronic) 2397-334X (Linking). doi:10.1038/s41559-020-1129-2.

G. Hanson and J. Coller. Codon optimality, bias and usage in translation and mrna decay. *Nat Rev Mol Cell Biol*, 19(1):20–30, 2018. ISSN 1471-0080 (Electronic) 1471-0072 (Print) 1471-0072 (Linking). doi:10.1038/nrm.2017.91.

Peter W. Harrison, Alison E. Wright, Fabian Zimmer, Rebecca Dean, Stephen H. Montgomery, Marie A. Pointer, and Judith E. Mank. Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences*, 112(14):4393–4398, 2015. doi:doi:10.1073/pnas.1501339112.

Tina M. Henkin and Charles Yanofsky. Regulation by transcription attenuation in bacteria: how rna provides instructions for transcription termination/antitermination decisions. *BioEssays*, 24(8):700–707, 2002. ISSN 0265-9247. doi:10.1002/bies.10125.

Chan PP Holmes AD, Howard JM and Lowe TM. trna analysis of expression (trax): A tool for integrating analysis of trnas, trna-derived small rnas, and trna modifications., 2020.

Richard Holmquist, Thomas H. Jukes, and Sharon Pangburn. Evolution of transfer rna. *Journal of Molecular Biology*, 78(1):91–116, 1973. ISSN 0022-2836. doi:https://doi.org/10.1016/0022-2836(73)90430-0.

Hans A Hosbach, Melvin Silberklang, and Brian J McCarthy. Evolution of a d. melanogaster glutamate trna gene cluster. *Cell*, 21(1):169–178, 1980. ISSN 0092-8674.

A. M. Hudson, N. L. Szabo, G. Loughran, N. M. Wills, J. F. Atkins, and L. Cooley. Tissue-specific dynamic codon redefinition in drosophila. *Proc Natl Acad Sci U S A*, 118(5), 2021. ISSN 1091-6490 (Electronic) 0027-8424 (Print) 0027-8424 (Linking). doi:10.1073/pnas.2012793118.

Richard R. Hudson. Generating samples under a wright–fisher neutral model of genetic variation. *Bioinformatics*, 18(2):337–338, 2002. ISSN 1367-4803. doi:10.1093/bioinformatics/18.2.337.

Richard R Hudson, Martin Kreitman, and Montserrat Aguadé. A test of neutral molecular evolution based on nucleotide data. *Genetics*, 116(1):153–159, 1987. ISSN 1943-2631. doi:10.1093/genetics/116.1.153.

L. A. Hughes, D. L. Rudler, S. J. Siira, T. McCubbin, S. A. Raven, J. M. Browne, J. A. Ermer, J. Rientjes, J. Rodger, E. Marcellin, O. Rackham, and A. Filipovska. Copy number variation in trna isodecoder genes impairs mammalian development and balanced translation. *Nat Commun*, 14(1):2210, 2023. ISSN 2041-1723 (Electronic) 2041-1723 (Linking). doi:10.1038/s41467-023-37843-9.

Toshimichi Ikemura. Codon usage and trna content in unicellular and multiceullular organisms. *Molecular Biology and Evolution*, 2(1):13–34, 1985.

P. Innocenti and E. H. Morrow. The sexually antagonistic genes of drosophila melanogaster. *PLoS Biol*, 8(3):e1000335, 2010. ISSN 1544-9173 (Print) 1544-9173. doi:10.1371/journal.pbio.1000335.

Dotu I Zhou H Yang XL Schimmel P Senju S Nishimura Y Chuang JH Ackerman SL. Ishimura R, Nagy G. Ribosome stalling induced by mutation of a cns-specific trna causes neurodegeneration. *Science*, 345(6195):455–9, 2014.

Fiona M Jucker and ARTHUR Pardi. Gnra tetraloops make a u-turn. *Rna*, 1(2):219, 1995.

P. Karki, T. D. Carney, C. Maracci, A. S. Yatsenko, H. R. Shcherbata, and M. V. Rodnina. Tissue-specific regulation of translational readthrough tunes functions of the traffic jam transcription factor. *Nucleic Acids Res*, 50(11):6001–6019, 2022. ISSN 1362-4962 (Electronic) 0305-1048 (Print) 0305-1048 (Linking). doi:10.1093/nar/gkab1189.

Bhavatharini Kasinathan, III Colmenares, Serafin U., Hannah McConnell, Janet M. Young, Gary H. Karpen, and Harmit S. Malik. Innovation of heterochromatin functions drives rapid evolution of essential zad-znf genes in drosophila. *eLife*, 9:e63368, 2020. ISSN 2050-084X. doi:10.7554/eLife.63368.

Simon P. Keam and Gyorgy Hutvagner. trna-derived fragments (trfs): Emerging new roles for an ancient rna in the regulation of gene expression. *Life*, 5(4):1638–1651, 2015. doi:10.3390/life5041638.

Motoo Kimura. *The neutral theory of molecular evolution*. Cambridge university press, 1985. ISBN 1139935674.

P. P. Kokate, S. M. Techtmann, and T. Werner. Codon usage bias and dinucleotide preference in 29 drosophila species. *G3 (Bethesda)*, 11(8), 2021. ISSN 2160-1836 (Electronic) 2160-1836 (Linking). doi:10.1093/g3journal/jkab191.

Andy T Kong, Felipe V Leprevost, Dmitry M Avtonomov, Dattatreya Mellacheruvu, and Alexey I Nesvizhskii. Msfragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nature methods*, 14(5):513–520, 2017. ISSN 1548-7091.

Srikar Krishna and et al. Dynamic expression of trna-derived small rnas define cellular states. *EMBO Rep*, 20(7):e47789, 2019. doi:10.15252/embr.201947789.

Thomas S. Kuhn. *The structure of scientific revolutions*. Second edition, enlarged. University of Chicago Press., 1970.

S. Kumar, G. Stecher, M. Suleski, and S. B. Hedges. Timetree: A resource for timelines, timetrees, and divergence times. *Mol Biol Evol*, 34(7):1812–1819, 2017. ISSN 1537-1719 (Electronic) 0737-4038 (Linking). doi:10.1093/molbev/msx116.

Cedric Landerer, Jonas Poehls, and Agnes Toth-Petroczy. Fitness effects of phenotypic mutations at proteome-scale reveal optimality of translation machinery. *Molecular biology and evolution*, 41(3):msae048, 2024a. ISSN 0737-4038.

Cedric Landerer, Maxim Scheremetjew, HongKee Moon, Lena Hersemann, and Agnes Toth-Petroczy. detelpy: Python package for high-throughput detection of amino acid substitutions in mass spectrometry datasets. *Bioinformatics*, 40(7), 2024b. ISSN 1367-4811. doi:10.1093/bioinformatics/btae424.

S. Ledoux, M. Olejniczak, and O. C. Uhlenbeck. A sequence element that tunes escherichia coli trna(ala)(ggc) to ensure accurate decoding. *Nat Struct Mol Biol*, 16

(4):359–64, 2009. ISSN 1545-9985 (Electronic) 1545-9993 (Print) 1545-9985 (Linking). doi:10.1038/nsmb.1581.

Chuan Li, Wenfeng Qian, Calum J. Maclean, and Jianzhi Zhang. The fitness landscape of a trna gene. *Science*, 352(6287):837–840, 2016. doi:doi:10.1126/science.aae0568.

Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, 25(14):1754–1760, 2009. ISSN 1367-4803. doi:10.1093/bioinformatics/btp324.

Yuchen Li and et al. Type ii trna cleavage by the antiviral factor slfn11 restricts hiv-1 infection. *Nat Struct Mol Biol*, 25(6):466–475, 2018. doi:10.1038/s41594-018-0062-0.

M. Long, N. W. VanKuren, S. Chen, and M. D. Vibranovski. New gene evolution: little did we know. *Annu Rev Genet*, 47:307–33, 2013. ISSN 1545-2948 (Electronic) 0066-4197 (Linking). doi:10.1146/annurev-genet-111212-133301.

B. Loppin, D. Lepetit, S. Dorus, P. Couble, and T. L. Karr. Origin and neofunctionalization of a drosophila paternal effect gene essential for zygote viability. *Curr Biol*, 15(2):87–93, 2005. ISSN 0960-9822 (Print) 0960-9822 (Linking). doi:10.1016/j.cub.2004.12.071.

Tien Lu and et al. trna introns: Presence, processing, and evolutionary roles in eukaryotes. *Genomics*, 105(1):1–7, 2015. doi:10.1016/j.ygeno.2014.10.002.

Shiqi Luo, Feng He, Junjie Luo, Shengqian Dou, Yirong Wang, Annan Guo, and Jian Lu. Drosophila tsrnas preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response. *Nucleic Acids Research*, 46(10):5250–5268, 2018. ISSN 0305-1048. doi:10.1093/nar/gky189.

M. A. Machnicka, A. Olchowik, H. Grosjean, and J. M. Bujnicki. Distribution and frequencies of post-transcriptional modifications in trnas. *RNA Biol*, 11(12):1619–29, 2014. ISSN 1547-6286 (Print) 1547-6286. doi:10.4161/15476286.2014.992273.

G. Martinez. trna-derived small rnas: New players in genome protection against retrotransposons. *RNA Biol*, 15(2):170–175, 2018. ISSN 1547-6286 (Print) 1547-6286. doi:10.1080/15476286.2017.1403000.

Glenn R. Masson. Towards a model of gcn2 activation. *Biochem Soc Trans*, 47(5):1481–1488, 2019. ISSN 0300-5127 (Print) 0300-5127. doi:10.1042/bst20190331.

Y. Mei, J. Yong, H. Liu, Y. Shi, J. Meinkoth, G. Dreyfuss, and X. Yang. trna binds to cytochrome c and inhibits caspase activation. *Mol Cell*, 37(5):668–78, 2010. ISSN 1097-4164 (Electronic) 1097-2765 (Print) 1097-2765 (Linking). doi:10.1016/j.molcel.2010.01.023.

Jacques Monod. *Chance and necessity*. Knopf, New York, 1971. Translated from the French edition (Paris, 1970) by Austryn Wainhouse.

Frank V. Murphy and V. Ramakrishnan. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nature Structural  Molecular Biology*, 11(12):1251–1252, 2004. ISSN 1545-9985. doi:10.1038/nsmb866.

E. P. Nawrocki and S. R. Eddy. Infernal 1.1: 100-fold faster rna homology searches. *Bioinformatics*, 29(22):2933–5, 2013. ISSN 1367-4803 (Print) 1367-4803. doi:10.1093/bioinformatics/btt509.

N. Netzer, J. M. Goodenbour, A. David, K. A. Dittmar, R. B. Jones, J. R. Schneider, D. Boone, E. M. Eves, M. R. Rosner, J. S. Gibbs, A. Embry, B. Dolan, S. Das, H. D. Hickman, P. Berglund, J. R. Bennink, J. W. Yewdell, and T. Pan. Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature*, 462(7272): 522–526, 2009. doi:10.1038/nature08576.

M. W. Nirenberg and J. H. Matthaei. The dependence of cell-free protein synthesis in e. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A*, 47 (10):1588–602, 1961. ISSN 0027-8424 (Print) 0027-8424. doi:10.1073/pnas.47.10.1588.

E. M. Novoa, M. Pavon-Eternod, T. Pan, and L. Ribas de Pouplana. A role for trna modifications in genome structure and codon usage. *Cell*, 149(1):202–13, 2012. ISSN 1097-4172 (Electronic) 0092-8674 (Linking). doi:10.1016/j.cell.2012.01.050.

D. J. Obbard, J. Maclennan, K. W. Kim, A. Rambaut, P. M. O'Grady, and F. M. Jiggins. Estimating divergence dates and substitution rates in the drosophila phylogeny. *Mol Biol Evol*, 29(11):3459–73, 2012. ISSN 1537-1719 (Electronic) 0737-4038 (Print) 0737-4038 (Linking). doi:10.1093/molbev/mss150.

Susumu Ohno. *Evolution by gene duplication*. Springer-Verlag, Berlin, 1970.

E. A. Orellana, E. Siegal, and R. I. Gregory. trna dysregulation and disease. *Nat Rev Genet*, 23(11):651–664, 2022. ISSN 1471-0064 (Electronic) 1471-0056 (Linking). doi:10.1038/s41576-022-00501-9.

Gayann I. Hall Otto Hagenbuchle, Drena Larson and Karen U. Sprague. The primary transcription product of a silkworm alanine trna gene: Identification of in vitro sites of initiation, termination and processing. *Cell*, 17:1217–1229, 1979.

Jian-Bo Pan, Shi-Chang Hu, Dan Shi, Mei-Chun Cai, Yin-Bo Li, Quan Zou, and Zhi-Liang Ji. Pagenbase: A pattern gene database for the global and dynamic understanding of gene function. *PLOS ONE*, 8(12):e80747, 2013. doi:10.1371/journal.pone.0080747.

T. Pan. Modifications and functional genomics of human transfer rna. *Cell Res*, 28 (4):395–404, 2018. ISSN 1748-7838 (Electronic) 1001-0602 (Print) 1001-0602 (Linking). doi:10.1038/s41422-018-0013-y.

M. Parisien, X. Wang, and T. Pan. Diversity of human trna genes from the 1000-genomes project. *RNA Biol*, 10(12):1853–67, 2013. ISSN 1555-8584 (Electronic) 1547-6286 (Print) 1547-6286 (Linking). doi:10.4161/rna.27361.

John Peden. codonw, v1.4.4. 2005.

O. Pinkard, S. McFarland, T. Sweet, and J. Coller. Quantitative trna-sequencing uncovers metazoan tissue-specific trna regulation. *Nat Commun*, 11(1):4104, 2020. ISSN 2041-1723 (Electronic) 2041-1723 (Linking). doi:10.1038/s41467-020-17879-x.

Ryan Poplin, Valentin Ruano-Rubio, Mark A. DePristo, Tim J. Fennell, Mauricio O. Carneiro, Geraldine A. Van der Auwera, David E. Kling, Laura D. Gauthier, Ami Levy-Moonshine, David Roazen, Khalid Shakir, Joel Thibault, Sheila Chandran, Chris Whelan, Monkol Lek, Stacey Gabriel, Mark J Daly, Ben Neale, Daniel G. MacArthur, and Eric Banks. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, page 201178, 2018. doi:10.1101/201178.

Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I. W. de Bakker, Mark J. Daly, and Pak C. Sham. Plink: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3):559–575, 2007. ISSN 0002-9297. doi:https://doi.org/10.1086/519795.

Wenfeng Qian and Jianzhi Zhang. Codon usage bias and nuclear mrna concentration: Correlation vs. causation. *Proceedings of the National Academy of Sciences*, 118(20): e2104714118, 2021. doi:doi:10.1073/pnas.2104714118.

Gary J. Quigley and Alexander Rich. Structural domains of transfer rna molecules. *Science*, 194(4267):796–806, 1976. doi:doi:10.1126/science.790568.

A. R. Quinlan and I. M. Hall. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–2, 2010. ISSN 1367-4811 (Electronic) 1367-4803 (Print) 1367-4803 (Linking). doi:10.1093/bioinformatics/btq033.

J. R. Raab, J. Chiu, J. Zhu, S. Katzman, S. Kurukuti, P. A. Wade, D. Haussler, and R. T. Kamakaka. Human trna genes function as chromatin insulators. *Embo j*, 31(2):330–50, 2012. ISSN 0261-4189 (Print) 0261-4189. doi:10.1038/emboj.2011.406.

Roni Rak, Orna Dahan, and Yitzhak Pilpel. Repertoires of trnas: The couplers of genomics and proteomics. *Annual Review of Cell and Developmental Biology*, 34(1):239–264, 2018. doi:10.1146/annurev-cellbio-100617-062754.

L. Randau and D. Soll. Transfer rna genes in pieces. *EMBO Rep*, 9(7):623–8, 2008. ISSN 1469-3178 (Electronic) 1469-221X (Print) 1469-221X (Linking). doi:10.1038/embor.2008.101.

Marie Sissler Richard Giege and Catherine Florentz. Universal rules and idiosyncratic features in trna identity. *Nucleic Acids Research*, 26(22):5017–5035, 1988.

H. H. Rogers, C. M. Bergman, and S. Griffiths-Jones. The evolution of trna genes in drosophila. *Genome Biol Evol*, 2:467–77, 2010. ISSN 1759-6653 (Electronic) 1759-6653 (Linking). doi:10.1093/gbe/evq034.

Rebekah L. Rogers, Julie M. Cridland, Ling Shao, Tina T. Hu, Peter Andolfatto, and Kevin R. Thornton. Landscape of standing variation for tandem duplications in drosophila yakuba and drosophila simulans. *Molecular Biology and Evolution*, 31(7):1750–1766, 2014. ISSN 0737-4038. doi:10.1093/molbev/msu124.

B. D. Ross, L. Rosin, A. W. Thomae, M. A. Hiatt, D. Vermaak, A. F. de la Cruz, A. Imhof, B. G. Mellone, and H. S. Malik. Stepwise evolution of essential centromere function in a drosophila neogene. *Science*, 340(6137):1211–4, 2013. ISSN 1095-9203 (Electronic) 0036-8075 (Linking). doi:10.1126/science.1234393.

Joëlle Rudinger-Thirion, Alain Lescure, Charlotte Paulus, and Magali Frugier. Misfolded human trna isodecoder binds and neutralizes a 3' utr-embedded alu element. *Proc Natl Acad Sci U S A*, 108(40):E794–E802, 2011. doi:10.1073/pnas.1103698108.

M. Saikia and M. Hatzoglou. Angiogenin-cleaved trna halves interact with cytochrome c, protecting cells from apoptosis during osmotic stress. *Mol Cell Biol*, 34(13):2450–2463, 2014. doi:10.1128/MCB.00136-14.

Martin Schlee and Gunther Hartmann. Discriminating self from non-self in nucleic acid sensing. *Nat Rev Immunol*, 16(9):566–580, 2016. doi:10.1038/nri.2016.78.

Andrea J. Schorn and et al. Ltr-retrotransposon control by trna-derived small rnas. *Cell*, 170(1):61–71.e11, 2017. doi:10.1016/j.cell.2017.06.013.

L. L. C. Schrödinger. The pymol molecular graphics system, version 3.0, 2015/11 2015. PyMOL The PyMOL Molecular Graphics System, Version 3.0, Schrödinger, LLC.

David B. Schuntermann and et al. The central role of transfer rnas in mistranslation. *J Biol Chem*, 300(3):105714, 2024. doi:10.1016/j.jbc.2024.105714.

S. Sekine, O. Nureki, K. Sakamoto, T. Niimi, M. Tateno, M. Go, T. Kohno, A. Brisson, J. Lapointe, and S. Yokoyama. Major identity determinants in the "augmented d helix" of trna(glu) from escherichia coli. *J Mol Biol*, 256(4):685–700, 1996. ISSN 0022-2836 (Print) 0022-2836. doi:10.1006/jmbi.1996.0118.

U. Sharma, C. C. Conine, J. M. Shea, A. Boskovic, A. G. Derr, X. Y. Bing, C. Belleannee, A. Kucukural, R. W. Serra, F. Sun, L. Song, B. R. Carone, E. P. Ricci, X. Z. Li, L. Fauquier, M. J. Moore, R. Sullivan, C. C. Mello, M. Garber, and O. J. Rando. Biogenesis and function of trna fragments during sperm maturation and fertilization in mammals. *Science*, 351(6271):391–396, 2016. ISSN 0036-8075 (Print) 0036-8075. doi:10.1126/science.aad6780.

Max J. Shurtleff, Morayma M. Temoche-Diaz, Kalliopi V. Karfilis, Randy Schekman, and Robert H. Scheller. Broad role for ybx1 in defining the small noncoding rna composition of exosomes. *Proc Natl Acad Sci U S A*, 114(42):E8987–E8995, 2017. doi:10.1073/pnas.1712108114.

D. Sosa, J. Chen, S. Xia, M. Sobzcyk, T. Pan, and M. Long. Data and code from: Sex-biased duplicates are rapidly generated during drosophila trna repertoire evolution. page zenodo.org/doi/10.5281/zenodo.10501205., 2025. doi:zenodo.org/doi/10.5281/zenodo.10501205.

Magali Soumillon, Anamaria Necsulea, Manuela Weier, David Brawand, Xiaolan Zhang, Hongcang Gu, Pauline Barthès, Maria Kokkinaki, Serge Nef, Andreas Gnirke, Martin Dym, Bernard de Massy, Tarjei S Mikkelsen, and Henrik Kaessmann. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Reports*, 3 (6):2179–2190, 2013. ISSN 2211-1247. doi:https://doi.org/10.1016/j.celrep.2013.05.031.

Todd M Lowe Steven J Marygold, Patricia P Chan. Systematic identification of trna genes in drosophila melanogaster. *microPublication Biology*, 2022.

Ryan Stikeleather, Farhan Ali, Wei-Chin Ho, Tim Licknack, and Michael Lynch. Translation accuracy in e. coli. *bioRxiv*, page 2025.04.18.649569, 2025. doi:10.1101/2025.04.18.649569.

Z. Su, B. Wilson, P. Kumar, and A. Dutta. Non-canonical roles of trnas: trna fragments and beyond. *Annu Rev Genet*, 54:47–69, 2020. ISSN 1545-2948 (Electronic) 0066-4197 (Print) 0066-4197 (Linking). doi:10.1146/annurev-genet-022620-101840.

Arvind R. Subramaniam and et al. Transcriptional activators in the bacterial world. *Prog Mol Biol Transl Sci*, 117:1–32, 2013. doi:10.1016/B978-0-12-386497-0.00001-0.

D. T. Tang, E. A. Glazov, S. M. McWilliam, W. C. Barris, and B. P. Dalrymple. Analysis of the complement and molecular evolution of trna genes in cow. *BMC Genomics*, 10:188, 2009. ISSN 1471-2164 (Electronic) 1471-2164 (Linking). doi:10.1186/1471-2164-10-188.

I. Targanski and V. Cherkasova. Analysis of genomic trna sets from bacteria, archaea, and eukarya points to anticodon-codon hydrogen bonds as a major determinant of trna compositional variations. *Rna*, 14(6):1095–109, 2008. ISSN 1355-8382 (Print) 1355-8382. doi:10.1261/rna.896108.

B. P. Thornlow, J. Hough, J. M. Roger, H. Gong, T. M. Lowe, and R. B. Corbett-Detig. Transfer rna genes experience exceptionally elevated mutation rates. *Proc Natl Acad Sci U S A*, 115(36):8996–9001, 2018. ISSN 1091-6490 (Electronic) 0027-8424 (Print) 0027-8424 (Linking). doi:10.1073/pnas.1801240115.

J. Tonzetich, S. Hayashi, and T. A. Grigliatti. Conservatism of sites of trna loci among the linkage groups of several drosophila species. *J Mol Evol*, 30(2):182–8, 1990. ISSN 0022-2844 (Print) 0022-2844. doi:10.1007/bf02099944.

Juan Pablo Tosar and et al. Dimerization confers increased stability to nucleases in 5' halves from glycine and glutamic acid trnas. *Nucleic Acids Res*, 46(17):9080–9093, 2018. doi:10.1093/nar/gky695.

Dan I. Andersson Ulfar Bergthorsson and John R. Roth. Ohno's dilemma: Evolution of new genes under continuous selection. *PNAS*, 104(43), 2007.

Kevin Van Bortle, Michael H. Nichols, Li Li, Chin-Tong Ong, Naomi Takenaka, Zhaohui S. Qin, and Victor G. Corces. Topological organization and dynamic regulation of human trna genes during macrophage differentiation. *Genome Biol*, 18:180, 2017. doi:10.1186/s13059-017-1310-3.

Bram J. van Raam and Guy S. Salvesen. Transferring death: A role for trna in apoptosis regulation. *Molecular Cell*, 37(5):591–592, 2010. ISSN 1097-2765. doi:https://doi.org/10.1016/j.molcel.2010.02.001.

N. W. VanKuren and M. Long. Gene duplicates resolving sexual conflict rapidly evolved essential gametogenesis functions. *Nat Ecol Evol*, 2(4):705–712, 2018. ISSN 2397-334X (Electronic) 2397-334X (Linking). doi:10.1038/s41559-018-0471-0.

M. D. Vibranovski, Y. Zhang, and M. Long. General gene movement off the x chromosome in the drosophila genus. *Genome Res*, 19(5):897–903, 2009. ISSN 1088-9051 (Print) 1549-5477 (Electronic) 1088-9051 (Linking). doi:10.1101/gr.088609.108.

Nicolas Vinckenbosch, Isabelle Dupanloup, and Henrik Kaessmann. Evolutionary fate of retroposed gene copies in the human genome. *Proceedings of the National Academy of Sciences*, 103(9):3220–3225, 2006. doi:doi:10.1073/pnas.0511307103.

Wenkai Wang, Chenjie Feng, Renmin Han, Ziyi Wang, Lisha Ye, Zongyang Du, Hong Wei, Fa Zhang, Zhenling Peng, and Jianyi Yang. trrosettarna: automated prediction of rna 3d structure with transformer network. *Nature Communications*, 14(1):7266, 2023a. ISSN 2041-1723. doi:10.1038/s41467-023-42528-4.

Y. Wang, P. McNeil, R. Abdulazeez, M. Pascual, S. E. Johnston, P. D. Keightley, and D. J. Obbard. Variation in mutation, recombination, and transposition rates in drosophila melanogaster and drosophila simulans. *Genome Res*, 33(4):587–598, 2023b. ISSN 1088-9051 (Print) 1088-9051. doi:10.1101/gr.277383.122.

C. P. Watkins, W. Zhang, A. C. Wylder, C. D. Katanski, and T. Pan. A multiplex platform for small rna sequencing elucidates multifaceted trna stress response and translational regulation. *Nat Commun*, 13(1):2491, 2022. ISSN 2041-1723 (Electronic) 2041-1723 (Linking). doi:10.1038/s41467-022-30261-3.

E. Westhof, B. Thornlow, P. P. Chan, and T. M. Lowe. Eukaryotic trna sequences present conserved and amino acid-specific structural signatures. *Nucleic Acids Res*, 50 (7):4100–4112, 2022. ISSN 1362-4962 (Electronic) 0305-1048 (Print) 0305-1048 (Linking). doi:10.1093/nar/gkac222.

Carrie A. Whittle and Cassandra G. Extavour. Evidence for non-optimal codon choice in highly transcribed sex-biased genes of drosophila melanogaster. *bioRxiv*, page 2024.11.24.625100, 2024. doi:10.1101/2024.11.24.625100.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, 2016. ISBN 978-3-319-24277-4. URL `https://ggplot2.tidyverse.org`.

E. Wiltrout, J. M. Goodenbour, M. Fréchin, and T. Pan. Misacylation of trna with methionine in saccharomyces cerevisiae. *Nucleic Acids Res*, 40(20):10494–506, 2012. ISSN 0305-1048 (Print) 0305-1048. doi:10.1093/nar/gks805.

Kim Sneppen Winther and Kenn Gerdes. Enteric virulence associated protein vapc inhibits translation by cleavage of initiator trna. *Proc Natl Acad Sci U S A*, 108(18):7403–7407, 2011. doi:10.1073/pnas.1019587108.

M. Withers, L. Wernisch, and M. dos Reis. Archaeology and evolution of transfer rna genes in the escherichia coli genome. *Rna*, 12(6):933–42, 2006. ISSN 1355-8382 (Print) 1355-8382. doi:10.1261/rna.2272306.

I. Wohlgemuth, C. Pohl, J. Mittelstaet, A. L. Konevega, and M. V. Rodnina. Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Philos Trans R Soc Lond B Biol Sci*, 366(1580):2979–86, 2011. ISSN 1471-2970 (Electronic) 0962-8436 (Print) 0962-8436 (Linking). doi:10.1098/rstb.2011.0138.

F. Wright. The 'effective number of codons' used in a gene. *Gene*, 87(1):23–9, 1990. ISSN 0378-1119 (Print) 0378-1119. doi:10.1016/0378-1119(90)90491-9.

S. Xia, N. W. VanKuren, C. Chen, L. Zhang, C. Kemkemer, Y. Shao, H. Jia, U. Lee, A. S. Advani, A. Gschwend, M. D. Vibranovski, S. Chen, Y. E. Zhang, and M. Long. Genomic analyses of new genes and their phenotypic effects reveal rapid evolution of essential functions in drosophila development. *PLoS Genetics*, 17(7):e1009654, 2021. ISSN 1553-7390 (Print) 1553-7390. doi:10.1371/journal.pgen.1009654.

Shengqian Xia, Jianhai Chen, Deanna Arsala, J. J. Emerson, and Manyuan Long. Functional innovation through new genes as a general evolutionary process. *Nature Genetics*, 57(2): 295–309, 2025. ISSN 1546-1718. doi:10.1038/s41588-024-02059-0.

Shigetada Yamasaki, Pavel Ivanov, Guo-Fu Hu, and Paul Anderson. Angiogenin cleaves trna and promotes stress-induced translational repression. *J Cell Biol*, 185(1):35–42, 2009. doi:10.1083/jcb.200811106.

Ziheng Yang. Paml 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8):1586–1591, 2007. ISSN 0737-4038. doi:10.1093/molbev/msm088.

A. H. Yona, Z. Bloom-Ackermann, I. Frumkin, V. Hanson-Smith, Y. Charpak-Amikam, Q. Feng, J. D. Boeke, O. Dahan, and Y. Pilpel. trna genes rapidly change in evolution to meet novel translational demands. *Elife*, 2:e01339, 2013. ISSN 2050-084X (Print) 2050-084X (Electronic) 2050-084X (Linking). doi:10.7554/eLife.01339.

Fengchao Yu, Sarah E Haynes, Guo Ci Teo, Dmitry M Avtonomov, Daniel A Polasky, and Alexey I Nesvizhskii. Fast quantitative analysis of timstof pasef data with msfragger and ionquant. *Molecular Cellular Proteomics*, 19(9):1575–1585, 2020. ISSN 1535-9476.

J. M. Zaborske, V. L. DuMont, E. W. Wallace, T. Pan, C. F. Aquadro, and D. A. Drummond. A nutrient-driven trna modification alters translational fidelity and genome-wide protein coding across an animal genus. *PLoS Biol*, 12(12):e1002015, 2014. ISSN 1545-7885 (Electronic) 1544-9173 (Print) 1544-9173 (Linking). doi:10.1371/journal.pbio.1002015.

C. Zhang, M. Shine, A. M. Pyle, and Y. Zhang. Us-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat Methods*, 19(9):1109–1115, 2022. ISSN 1548-7091. doi:10.1038/s41592-022-01585-1.

J. Zhang, M. Long, and L. Li. Translational effects of differential codon usage among intragenic domains of new genes in drosophila. *Biochim Biophys Acta*, 1728(3):135–42, 2005. ISSN 0006-3002 (Print) 0006-3002 (Linking). doi:10.1016/j.bbaexp.2005.02.004.

J. Y. Zhang and Q. Zhou. On the regulatory evolution of new genes throughout their life history. *Mol Biol Evol*, 36(1):15–27, 2019. ISSN 1537-1719 (Electronic) 0737-4038 (Linking). doi:10.1093/molbev/msy206.

L. Zhang, Y. Ren, T. Yang, G. Li, J. Chen, A. R. Gschwend, Y. Yu, G. Hou, J. Zi, R. Zhou, B. Wen, J. Zhang, K. Chougule, M. Wang, D. Copetti, Z. Peng, C. Zhang, Y. Zhang, Y. Ouyang, R. A. Wing, S. Liu, and M. Long. Rapid evolution of protein diversity by de novo origination in oryza. *Nat Ecol Evol*, 3(4):679–690, 2019. ISSN 2397-334X (Electronic) 2397-334X (Linking). doi:10.1038/s41559-019-0822-5.

Xuedi Zhang, Ju Peng, Menghua Wu, Angyang Sun, Xiangyu Wu, Jie Zheng, Wangfei Shi, and Guanjun Gao. Broad phosphorylation mediated by testis-specific serine/threonine kinases contributes to spermiogenesis and male fertility. *Nature Communications*, 14(1): 2629, 2023. ISSN 2041-1723. doi:10.1038/s41467-023-38357-0.

Y. E. Zhang, M. D. Vibranovski, B. H. Krinsky, and M. Long. Age-dependent chromosomal distribution of male-biased genes in drosophila. *Genome Res*, 20(11):1526–33, 2010a. ISSN 1549-5469 (Electronic) 1088-9051 (Print) 1088-9051 (Linking). doi:10.1101/gr.107334.110.

Y. E. Zhang, M. D. Vibranovski, P. Landback, G. A. Marais, and M. Long. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the x chromosome. *PLoS Biol*, 8(10), 2010b. ISSN 1545-7885 (Electronic) 1544-9173 (Print) 1544-9173 (Linking). doi:10.1371/journal.pbio.1000494.

Y. E. Zhang, P. Landback, M. D. Vibranovski, and M. Long. Accelerated recruitment of new brain development genes into the human genome. *PLoS Biol*, 9(10):e1001179, 2011. ISSN 1545-7885 (Electronic) 1544-9173 (Linking). doi:10.1371/journal.pbio.1001179.

Yunfang Zhang, Xudong Zhang, Jianqiang Shi, Luciana C. De Oliveira, Guihua Feng, Qinghua Shi, Jian Chen, Jing Qian, Ling Zeng, Jia Yu, Qi Zhou, and Qinghua Shi. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding rnas. *Nat Cell Biol*, 20(5):535–540, 2018. doi:10.1038/s41556-018-0087-6.

Q. Zhou, G. Zhang, Y. Zhang, S. Xu, R. Zhao, Z. Zhan, X. Li, Y. Ding, S. Yang, and W. Wang. On the origin of new genes in drosophila. *Genome Res*, 18(9):1446–55, 2008. ISSN 1088-9051 (Print) 1088-9051. doi:10.1101/gr.076588.108.

Xiaoqiang Zhu, Victor Emmanuel Cruz, He Zhang, Jan P. Erzberger, and Joshua T. Mendell. Specific trnas promote mrna decay by recruiting the ccr4-not complex to translating ribosomes. *Science*, 386(6724):eadq8587, 2024. doi:doi:10.1126/science.adq8587.

# Brief Curriculum Vitæ

The author was born in Victoria, Texas, U.S.A. during the '90s to Sonya Barrientes and Joe Sosa. He completed his primary education at Smith Elementary School, Howell Middle School, and Victoria East High School. It was in those halls where the author developed a life-long zeal for music performance and the fine arts. The author earned a B.S. in Bioinformatics from St. Edward's University under the mentorship of Professor Charles Hauser. He then earned a M.S. in Bioinformatics & Computational Biology from Saint Louis University where he was advised by Professor Dapeng Zhang. From the University of Chicago, the author graduates as a Doctor of Philosophy where he spent several happy years studying and researching alongside Professor Manyuan Long in the Department of Ecology & Evolution. In 2025, the author won the the Graduate Student Excellence Award from the Society for Molecular Biology & Evolution. As a Ph.D. candidate, the author gave selected or invited seminars on his doctoral work six times in Asia, Europe, and North America.

*Chicago*

*February, 2026* D. Sosa