Nuclear-encoded tRNA genes harbor substantial allelic diversity in three nematode species

Avery Davis Bell[1], Corinne Simonti[1,3], Hector Baños[2,4], Ling Wang[1], Christine Heitsch[2], Joseph Lachance[1], Annalise B. Paaby[1]*

[1] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA
[2] School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA
[3]Current address: AAAS, Washington, DC
[4]Current address: Department of Mathematics, California State University San Bernardino, San Bernardino, CA, USA
*Correspondence: paaby@gatech.edu

Abstract

Cytosolic transfer RNAs, which are encoded as hundreds of genes in eukaryotic nuclear genomes, experience exceptionally high rates of mutation and have been hypothesized to carry significant mutational load in natural populations. Comparisons across species show pervasive losses, gains, and functional remolding at tRNA loci, indicating rapid evolution over intermediate timescales, but the underlying molecular changes happening over short timescales remain poorly understood. A central gap in our understanding is the nature of deleterious tRNA mutations, how they affect fitness, and the extent to which tRNA pools vary functionally across genomes. Furthermore, new technologies for capturing tRNA expression and modification emphasize the importance of tRNA activity to organismal fitness, as tRNA dysregulation is associated with disease burden in humans, but this research area has not yet included possible consequences of mutational variation. To infer the dynamics of tRNA evolution over short timescales, and to integrate our understanding of fitness consequences of tRNA regulation with mutation load, we first require a comprehensive view of standing variation in cytosolic tRNAs within populations. In this study, we resolve within-species allelic variation in nuclear-encoded tRNAs in three nematode species: *Caenorhabditis elegans*, *C. briggsae*, and *C. tropicalis*. We show that these tRNA repertoires have been shaped by transcription-associated mutagenesis and selection and exhibit extensive allelic variation, including isotype switching between tRNA backbones and anticodons, and genomic organization and patterns of variation that differ markedly from those of protein-coding regions. Individual genomes harbor remarkably different pools of tRNA genes predicting a wide range tRNA functional complements within a species.

Introduction

Transfer RNAs are essential molecules that deliver individual amino acids to the growing polypeptide chain during protein synthesis. Universal across the tree of life, tRNAs are encoded as genes in nuclear and organelle genomes and exhibit striking conservation: sequences specify the canonical 'cloverleaf' secondary structure (**Figure 1A**) and L-shaped tertiary structure integral to their molecular function. However, despite such ancient conservation, tRNAs encoded in nuclear genomes also undergo extensive evolutionary turnover. Across eukaryotic lineages, tRNA repertoires exhibit copy number variation, reassignment of tRNA identity from one cognate amino acid to another (isotype switching), and frequent gains and losses, including by pseudogenization following accumulation of mutations (Bermudez-Santana et al. 2010; Rogers et al. 2010; Iben and Maraia 2012; Wang and Ruvinsky 2012; Rogers and Griffiths-Jones 2014; Velandia-Huerto et al. 2016). The fraction of tRNA loci that share syntenic orthology among extant primate species is estimated to be between one third and one half (Velandia-Huerto et al. 2016); between the species *Caenorhabditis elegans* and *C. briggsae*, also about one half (Wang and Ruvinsky 2012).

The evolutionary dynamism of tRNAs is likely driven by the strong, opposing evolutionary forces they experience. tRNAs are among the most highly transcribed genes in the genome and consequently are subject to extremely high rates of transcription-associated mutagenesis (TAM) (Thornlow et al. 2018). Like other aspects of tRNAs, this phenomenon appears universal: tRNA loci in humans, *Mus musculus*, *Drosophila melanogaster*, and *Arabidopsis thaliana* show strong signatures of TAM (Thornlow et al. 2018). Modeling the TAM signature via a sequence substitution model, we also see evidence of TAM in *C. elegans* tRNAs (Banos et al. 2025). This elevated mutational input at tRNA genes is opposed by strong purifying selection, evidenced by the observation that tRNA gene bodies exhibit dramatically lower mutational variation than their adjacent flanking regions; both components are exposed to TAM during transcription, but selection purges much of the variation on the functional sequence (Thornlow et al. 2018).

Despite this clear evidence of very strong purifying selection at tRNAs, the elevated rate of TAM on nuclear-encoded tRNAs may confer a significant mutation load on eukaryotic genomes (Thornlow et al. 2018). Exploring this hypothesis requires a clearer understanding of the fitness consequences of standing genetic variation in tRNAs within extant populations. New approaches to analyzing tRNA expression have shown that regulatory control of tRNAs is complex and dynamic across age, cell type, and environmental cues, and that tRNA dysregulation is associated with disease burden in humans (Pinzaru and Tavazoie 2023; Suzuki 2021). These findings directly link tRNA activity to organismal fitness, though this burgeoning research area has largely overlooked the consequence of mutational variation on post-transcriptional tRNA function. Mutations to mitochondrial tRNAs have long been known to cause an array of diseases (Kobayashi et al. 1990; Goto et al. 1990; Florentz et al. 2003; Yarham et al. 2010; Suzuki et al. 2011), but the occurrence of mt-tRNAs as typically single-copy genes makes them both more vulnerable to impacts of mutation and more tractable to study.

To resolve the micro-evolutionary forces acting on nuclear-encoded tRNAs in contemporaneous populations, and to integrate our understanding of the fitness consequences of tRNA regulation

with mutational variation, we need a comprehensive characterization of tRNA allelic variation at the population level. One reason that this remains an underexplored area may be the large number of tRNAs in nuclear genomes. Unlike mt-tRNAs, which are typically encoded as one gene per amino acid (22 total: 20 canonical plus two modified), cytosolic tRNAs in eukaryotic nuclear genomes are encoded in the hundreds, with often dozens per amino acid (Goodenbour and Pan 2006; Bermudez-Santana et al. 2010). This organization poses challenges for comparisons within populations as well as across lineages. These latter comparisons often require resolution of orthology that is often both biologically absent (due to genuine gene gains and losses) and difficult to infer syntenically (Lant et al. 2019; Velandia-Huerto et al. 2016; Wang and Ruvinsky 2012) .

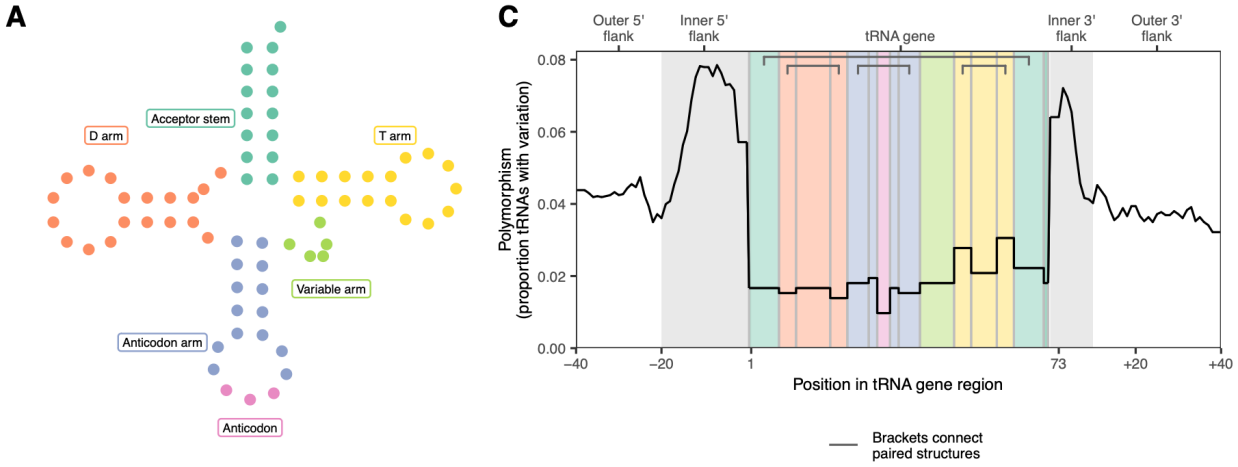
Nevertheless, interest is growing in the extent and functional impact of allelic variation in cytosolic tRNAs, especially in the context of human disease (Schoenmakers et al. 2016; Thornlow et al. 2018; Lant et al. 2019; Pinzaru and Tavazoie 2023; Murillo-Recio et al. 2025) . There is a consensus that the full extent of cytosolic tRNA variation remains underestimated (Lant et al. 2019). In humans, studies with small sample sizes likely under-sample the breadth of variation, yet they still reveal striking variation at tRNA genes (Iben and Maraia 2014; Darrow and Chadwick 2014; Berg et al. 2019). Larger-scale, lower-coverage sequencing also suggests extensive variation (Parisien et al. 2013; Lant et al. 2019), though these approaches do not capture the rare variants that likely account for much of functionally relevant allelic diversity. Even so, evidence is mounting that human tRNA mutations predicted to cause mis-loading of specific amino acids induce biochemical consequences (Davey-Young et al. 2024; Tennakoon et al. 2025).

The human genome encodes over 400 putatively functional cytosolic tRNA genes, which exhibit both identical and diverged sequence characteristics and diversity in the number of genes within each amino acid (isoacceptor) class (Goodenbour and Pan 2006; Chan and Lowe 2016). This complexity in the tRNA pool has been hypothesized to be adaptive, evolving to meet the proteomic demands of humans as complex, multicellular organisms (e.g., (Goodenbour and Pan 2006; Dittmar et al. 2006)). Yet the nematode worm *C. elegans*, approximately 30 billion times smaller than a human at ~1,000 cells (Sulston and Horvitz 1977), encodes over 500 tRNA genes and exhibits a similarly complex distribution of isoacceptors (Duret 2000). This similarity suggests that the functional contributions of tRNAs and the evolutionary pressures shaping tRNA pools are broadly conserved across eukaryotes. Cytosolic tRNAs of other eukaryotic systems likewise number in the hundreds with variable isoacceptor classes; in yeast, plants, and mice, intraspecific investigations indicate copy number variation and evidence of genetic context-specific functional consequences (Iben and Maraia 2012; Ishimura et al. 2014; Szymanski et al. 2023), despite continued under-ascertainment of the breadth of intraspecific variation.

The universality of tRNAs, including their essentiality, organization in eukaryotic nuclear genomes, and vulnerability to TAM, motivate comprehensive, high-resolution characterization of their allelic variation—in any representative system. *Caenorhabditis* nematodes offer an ideal opportunity. The model organism *C. elegans* is supported by extensive resources, including hundreds of genetically diverse strains and associated quantitative genetics tools, maximizing

tractability in follow-on experiments (Barriere and Felix 2005a; Barriere and Felix 2005b; Félix et al. 2013; Crombie et al. 2019; Andersen and Rockman 2022; Crombie et al. 2024). *C. elegans* belongs to the well-characterized *Caenorhabditis* phylogeny that includes *C. briggsae* and *C. tropicalis*, species with concomitant collections of sequenced strains (Crombie et al. 2024). These three species diverged from each other ~10-25 million years ago, with ~20-30% synonymous site differences (Stevens et al. 2022; Fusca et al. 2025). All three species are androdioecious, *i.e.*, hermaphroditic and largely self-fertilizing in nature; this mating system may promote the persistence of deleterious variants due to reduced recombination and less efficient selection (Barriere and Felix 2005a; Dolgin et al. 2007; Loewe and Cutter 2008; Frézal and Félix 2015; Fusca et al. 2025). Thus, deleterious changes to tRNAs may be more easily captured in population samples of these species. Comparison of tRNA repertoires across species also enables inferences about the timescale of observed mutations and whether they represent species-specific changes or broader evolutionary patterns. These species also differ in levels of within-species genetic diversity, providing an additional axis for comparative analysis.

In this study, we took advantage of >620 wild strains in each of three *Caenorhabditis* species (Crombie et al. 2024) to resolve within-species allelic variation in nuclear-encoded tRNAs. We show that these tRNA repertoires have been shaped by transcription-associated mutagenesis and selection and exhibit (i) extensive allelic variation, including isotype switching between tRNA backbones (scaffolds) and anticodons, and (ii) genomic organization and patterns of variation that differ markedly from those of protein-coding regions.



B

	What is invariant within group?	What may differ within group?	*C. elegans* N unique in reference genome	*C. briggsae* N unique in reference genome	*C. tropicalis* N unique in reference genome
Isotype (isoacceptor)	Cognate amino acid (AA)	Anticodon Sequence (outside of AA–specifying regions)	22	22	22
Isodecoder	Anticodon (cognate AA)	Sequence (outside of AA–specifying regions)	48	49	49
Sequence	Full gene sequence (isodecoder, cognate AA)	Genomic location	202	244	149
Gene	Genomic location, full gene sequence (isodecoder, cognate AA)	– nothing, single gene –	620	758	533

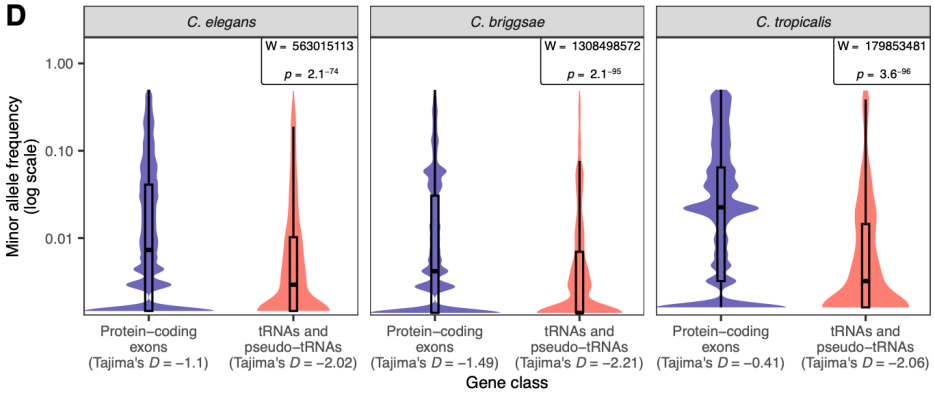


Figure 1. Structural, functional, sequence similarity, and transcription-associated mutagenesis (TAM) of tRNA genes in *Caenorhabditis* nematodes. **A,** "Cloverleaf" secondary structure of a canonical 72 bp tRNA, with each base plotted as a circle and colored and labeled by the component of the secondary structure it is part of (linker bases included with the neighboring arm). **B,** The number of tRNA genes in the reference genomes of three *Caenorhabditis* species corresponding to distinct amino acids (isotypes/isoacceptors), distinct anticodons (isodecoders), distinct full-tRNA sequence, and individual genes. **C**, Variation across tRNA gene regions in *C. elegans* for all tRNA genes in the reference genome combined (n = 720, includes all tRNAscan-SE predicted tRNAs including pseudogenes): the y axis is the proportion of tRNAs with mutational variation (variants with minor allele frequency < 0.05) observed in the 687-strain population at that base (5 bp moving window average). Regions in the tRNA gene itself are color coded according to secondary structure as in **A** while flanking regions showing the influence of TAM without selection are in white and gray. (Linker bases excluded.) **D,** Minor allele frequency (MAF)

spectrum of variants (SNVs and INDELs) in protein-coding exons and tRNA genes and tRNA gene-like sequences; between-gene set Mann-Whitney test results (top right) and Tajima's *D* values are annotated. Boxplots overlaid to show median (bold center bar), 25th and 75th data percentiles (box), and 1.5x interquartile range (whiskers). *N* variants per class per species in Table S3.

Results

tRNA repertoires in *Caenorhabditis*

The highly conserved 'cloverleaf' secondary structure and corresponding genomic sequence motifs of tRNAs (**Figure 1A**) enable computational prediction of the number and location of all predicted functional tRNAs and predicted non-functional tRNA-like sequences (pseudogenes) in a genome (Chan et al. 2021). Each tRNA gene belongs to multiple nested classes (**Figure 1B**): the gene itself with its unique location and sequence, a set of one or more genes with identical sequence but different genomic location, an isodecoder class of genes that share the same anticodon and therefore cognate amino acid (isodecoder), and a isotype (also called isoacceptor) class that shares the same cognate amino acid, with or without sharing the anticodon. To begin to understand tRNA evolution and variation in *Caenorhabditis* nematodes, we first generated a *Caenorhabditis* tRNA atlas of all nuclear-encoded (cytosolic) tRNAs and tRNA-like-sequences (pseudogenes) in the reference genomes of three species: model system *C. elegans*, and related *Caenorhabditis* species *C. briggsae* and *C. tropicalis.* Each species encoded tRNAs for all 20 canonical amino acids as well as initiation-specific methionine (iMet) and selenocysteine (SeC) from 48-49 distinct anticodon sequences; the nuclear reference genomes contained 149-244 unique tRNA gene sequences (**Figure 1B**, **Table S1-2**). Many of these sequences occurred at multiple genomic locations, resulting in 533-758 total tRNA location- and sequence-specific genes (**Figure 1B**). These patterns were consistent with known tRNA gene complements (Chan and Lowe 2009, 2016).

We first investigated isotype and isodecoder frequency at tRNA genes in these three species. Species showed broadly similar patterns (Pearson's $R^2 \geq 0.92$, $p < 4.3 \times 10^{-13}$ for gene number across isotypes), but also substantial copy number variation, with some isotypes differing by a third or more in gene count across species (**Table S1**). In all three species, selenocysteine (SeC) was the rarest tRNA encoded amino acid (**Table S1**), consistent with SeC's uniqueness across the tree of life (Chan and Lowe 2016; Commans and Bock 1999). In *C. elegans* and *C. briggsae*, glycine was the most common isotype (52 and 80 genes respectively; *C. tropicalis* has 50), while in *C. tropicalis* leucine was the most prevalent isotype (54 genes); *C. elegans* and *C. briggsae* also had many leucine tRNA genes (50 and 78, respectively) (**Table S1**).

At the codon level, in *C. elegans,* three codons were observed only once in tRNA genes (GCG/arginine, ACA/cysteine, and TCA/selenocysteine); in *C. briggsae*, five codons appeared in only one tRNA gene (GGC/ala, ATT/asparagine, ACA/cysteine, AAA/phenylalanine, and GGG, proline); and in *C. tropicalis*, two (CCG/arginine and TCA/selenocysteine) (**Table S2**). The most frequently observed codons were TCC/glycine in *C. elegans* and *C. briggsae* (36 and 60 genes, respectively) and TGG/proline in *C. tropicalis* (38 genes) (**Table S2**).

Caenorhabditis tRNA genes exhibit signatures of transcription-associated mutagenesis

tRNAs show evidence of exceptionally high rates of mutation, incurred because of their high rate of transcription (Thornlow et al. 2018). To assess the role of transcription-associated mutagenesis (TAM) in our system, we examined population variation at tRNA loci and their flanking genomic regions in our three species. tRNA gene flanking regions are unwound and vulnerable to mutation during transcription, but experience less selective pressure than do the tRNA gene body regions (Jinks-Robertson and Bhagwat 2014; Thornlow et al. 2018). Consistent with our earlier findings in *C. elegans* (Banos et al. 2025), tRNA genes and especially their flanking regions exhibited substantial TAM-like genetic variation across wild strains in all three species (**Figure 1C, S1, S2**). Specifically, tRNA gene flanking regions had exceptionally high mutational variation, with up to 7.6-12.9% of tRNA genes exhibiting mutations at a given base across species (**Figure 1C, Figure S1**). The more often TAM-induced mutations G>A and C>T (Gaillard and Aguilera 2016; Jinks-Robertson and Bhagwat 2014) tended to be the most frequent specific mutations in upstream tRNA gene flanking regions and within tRNA gene bodies (**Figure S2**). Regions coding for the tRNA cloverleaf itself had much less frequent mutations than flanking regions, evidencing selection's purging of tRNA mutations (**Figure S1**). Nevertheless, we observed mutations in all parts of the tRNA in each of the three species (**Figure S1**).

Given TAM's introduction of mutations and selection's assumed purging of these at tRNAs, we imagined that mutations at tRNAs might segregate in the population at lower frequencies than variants in other functionally important regions of the genome such as protein-coding regions. Indeed, variants in tRNA genes tended to have lower minor allele frequencies (**Figure 1D**) and these variants were much more likely to be present in the population as singletons than variants in protein-coding regions (median 29% singletons in tRNAs vs. 46% in protein-coding regions across species, **Table S3**). Commensurately, Tajima's *D* (Tajima 1989), which compares the number of pairwise differences between sequences and the number of segregating sites, was much lower in tRNAs than protein-coding regions, reflecting the excess of rare alleles at tRNAs even over protein-coding genes (**Figure 1D, Table S3**). These results point to the extreme mutation-selection balance shaping the evolution of tRNA genes in these species.

Allelic variation at *Caenorhabditis tRNA genes with putatively functional consequences*

We next sought to understand how TAM-derived variation manifests as allelic differences across *Caenorhabditis* tRNAs to address the question of how tRNA genes vary within populations. To do this, we focused on the set of tRNA genes present in the reference genome and reconstructed full-length alleles for each gene in each wild strain (>620 per species) by integrating variant genotypes from VCF files (Crombie et al. 2024). Across all nuclear-encoded tRNAs, we found 1695 distinct tRNA alleles in *C. elegans*, 3574 in *C. briggsae*, and 1048 in *C. tropicalis* (including predicted pseudogenized/non-functional alleles) (**Figure 2**). Individual tRNA isotypes comprised 1-139 alleles in *C. elegans*, 4-328 in *C. briggsae*, and 1-99 in *C. tropicalis*, while individual tRNA genes harbored between one and 17 alleles in *Caenorhabditis* populations (**Figure 2, Figure S3**). At putatively functional tRNA genes, 38.8-83.3% had allelic variation across the three species (**Figure S3**).

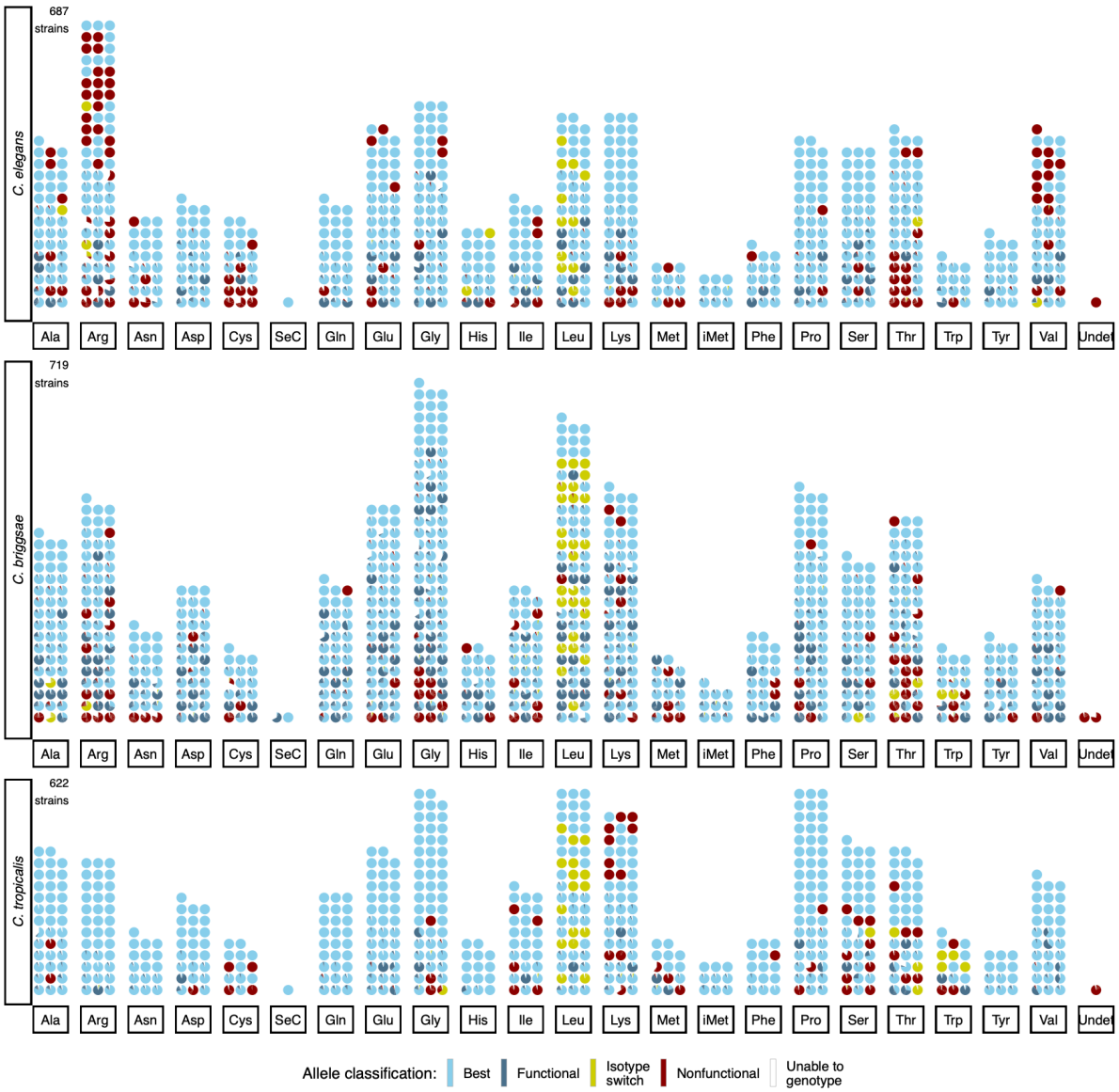


Figure 2. Allelic diversity at all tRNA genes in *Caenorhabditis* reference genomes. Each tRNA gene in the reference genome of each species is represented by a circle, grouped by the amino acid (isotype) predicted to match its backbone; slices of the circle correspond to different alleles of that gene (alleles with frequency < 0.05 plotted as having frequency = 0.05). Allele slices are colored by inferred functional classification (main text); white slices correspond to strains where one or more variants in that gene had missing genotypes. Genes may have multiple alleles of the same classification (color) but may have one and only one 'best' allele. The number of strains for which allelic data was generated is in the top left corner. Details on each gene, including zooms of each pie chart, are available via searchable web app at https://wildworm.biosci.gatech.edu/trnarefalleles/.

This remarkable allelic diversity, combined with high number of tRNA genes, resulted in each strain in *C. elegans* and *C. briggsae* having a unique complement of tRNA alleles genome-wide (including pseudogenes and missing-genotype genes). In *C. tropicalis*, which has fewer genotyped strains and less average genetic diversity (Noble et al. 2021), 94% of strains (545 of

577) had unique tRNA allelic complements, while the remaining strains shared tRNA allele identity with one (25 allelic combination), two (6 allelic combination), or eight (one allelic combination) other strains.

To begin to investigate potential functional relevance of this extensive allelic variation, we developed allele classifications by combining information about: the amino acid predicted from the tRNA backbone and evolutionary relatedness, excluding the anticodon ('backbone'); the anticodon's cognate amino acid; and the predicted tRNA folding score (all from tRNAscan-SE). Alleles where the anticodon and backbone were predicted to match the same amino acid were classified as 'Best' if they had the highest RNA folding score or 'Functional' if they had a putatively functional score lower than that of the 'Best' allele; in practice any of the 'Functional' alleles might truly be most functional. Alleles with nonfunctional folding scores or non-identifiable anticodons were classified as 'Nonfunctional.' Alleles where the backbone was predicted to match a different amino acid than the anticodon's cognate were classified as 'Isotype switch' alleles; isotype switching has long been observed across tRNA evolution at interspecies timescales (Rogers and Griffiths-Jones 2014).

While many nuclear tRNA genes harbored only one functional allele (obligately classified as 'Best'), many exhibited functional allelic diversity, with alleles belonging to all categories observed in all three species (**Figure 2**). While nonfunctional alleles naturally sometimes comprised the entirety of a tRNA pseudogene's alleles, in many cases a minority of alleles at a given gene were nonfunctional, suggesting that some strains in the population carry working tRNAs for that specific gene while others do not (**Figure 2**). Different tRNA isotypes had different predicted levels of functional allelic diversity, but this did not arise solely from the number of genes at a given isotype; for example, methionine has a low number of encoding tRNAs but a high number with predicted nonfunctional alleles (**Figure 2**).

Alleles with putative isotype switches would generally be predicted to be deleterious, as tRNAs encoded by those alleles might improperly load amino acids (Giege et al. 1998). Correspondingly, many isotype switching alleles occurred at very low frequency in the three *Caenorhabditis* species (**Figure 2**, small yellow pie pieces). On the other hand, some isotype switching may turn out to be beneficial or neutral and persist across taxa (Rogers and Griffiths-Jones 2014; Yona et al. 2013), a phenomenon that might manifest as the high-frequency, perhaps ancient isotype switches we observe (**Figure 2**, fully or nearly fully yellow pies). We next investigated these hypotheses further.

Evidence for both contemporaneous and ancient isotype switching in *Caenorhabditis*

To determine whether isotype switching occurs as polymorphic variation within extant populations and detect any changes that arose in a common ancestor of one or more *Caenorhabditis* species (~25 mya), we first assessed confidence in classifying an allele as an isotype switch. In the simplest case, tRNAscan-SE infers with high confidence that the allele's backbone corresponds to a different amino acid than its ancestral codon, an anticodon-based isotype switch. In other cases, the codon remains identical across alleles (or genes) but tRNAscan-SE infers different backbones, suggesting that mutations in non-anticodon regions underpin the switch, as has can occur via even a single nucleotide change (Hipps et al. 1995).

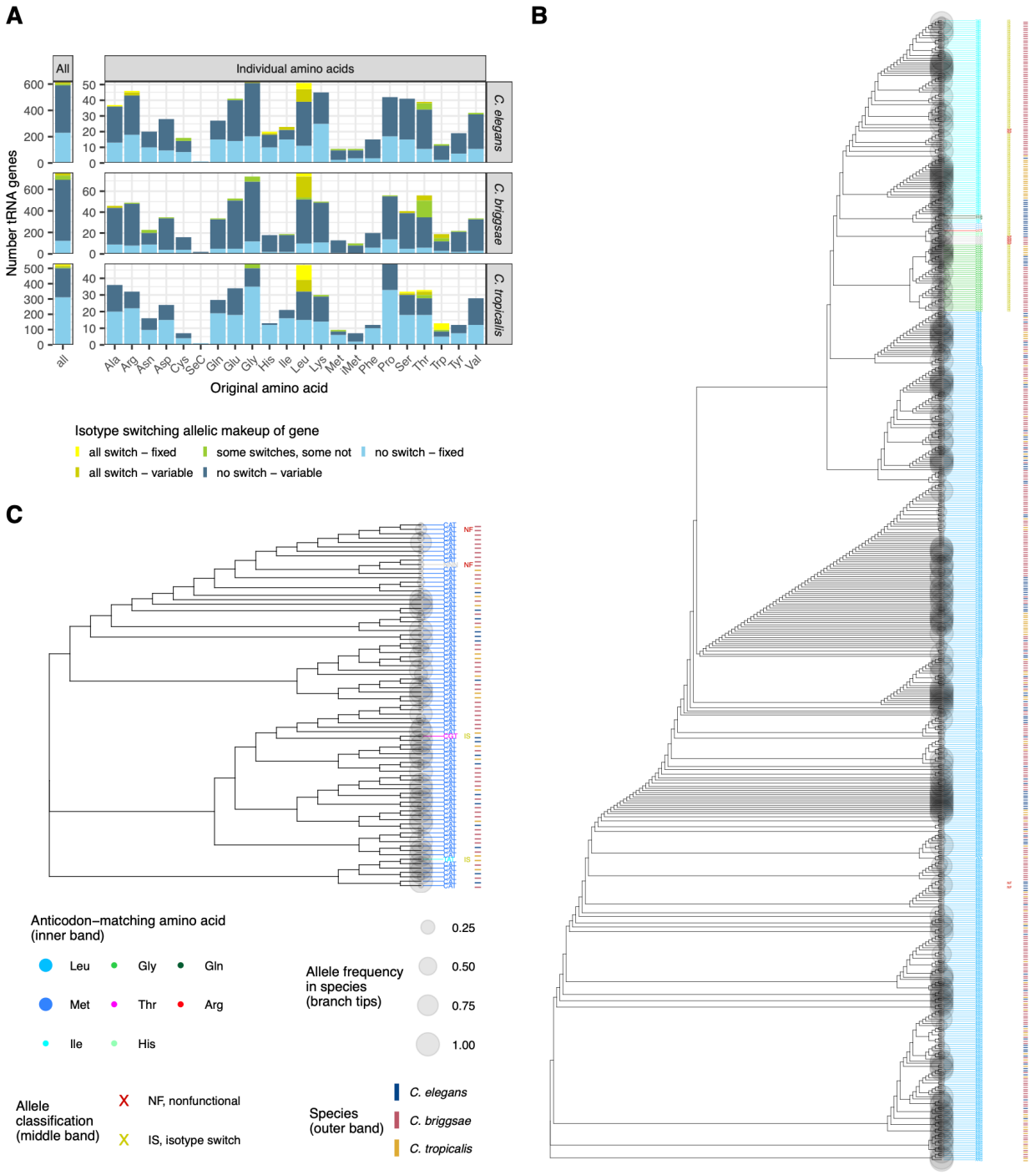


Figure 3. Isotype switching from ancient and contemporaneous tRNA genetic variation. **A,** The number of tRNA genes predicted to encode each amino acid colored by whether the gene's anticodon matches the prediction and whether there is allelic variation at that gene in the species (match: no switch; mismatch for all alleles: all switch; population varies in isotype switching or not: some switches, some not). **B** and **C**, neighbor-joining allele cladograms from whole-gene secondary-structure aware tRNA multiple sequence alignments (Methods) for all alleles from all species for genes predicted to have a leucine (**B**) or methionine (not iMet) (**C**) backbone. Branch label color reflects the amino acid encoded by the anticodon, with isotype switches noted in the central circle as yellow and predicted non-functional alleles noted in the

central circle as red. In **B**, ancestral leucines are now predicted to [match] isoleucine (cyan) in all strains in related alleles in all three species and glycine (green) in *C. briggsae* and *C. elegans*. In **C**, ancestral methionine now match threonine (magenta) in a minority of strains in *C. elegans* and isoleucine (cyan) in a minority of strains in *C. briggsae.*

When isotype switching was variable in the population, we could clearly map observed variants to the predicted isotype switch in the majority of cases, such that we continued with this analysis (**Figure S4,** Methods)

We characterized isotype switching in terms of 1) the frequency of isotype switching from each amino acid and 2) the frequency of the isotype switch allele(s) themselves (**Figure 3A**). For the latter, genes with isotype switching alleles fell into three major categories: 1) 'all switch': all alleles were isotype switches (either one or multiple alleles with the same switch); 2) 'No switch': all alleles retained the original isotype; and 3: 'Some switch, some not': genes with a mix of switching and non-switching alleles (**Figure 3A**). Fixed isotype switching occurred from leucine and threonine in all three species; from alanine, serine, and tryptophan in two of three species; and from arginine, glycine, histidine, and isoleucine in species-specific cases (**Figure 3A**, **Table S4**). The presence of isotype switching was not solely dependent on the number of genes per isotype. For example, tryptophan had among the fewest genes per isotype but showed fixed or variable isotype switching in all three species, while proline had among the highest gene count but only exhibited low-frequency variable switching in *C. briggsae* (**Figure 3A**). The amount of fixed isotype switching, presumably arising in a common ancestor of all strains in a species, did not correlate with overall genetic or tRNA allelic variation: *C. elegans* (intermediate variation) had 7 amino acids with fixed switching, *C. briggsae* (highest variation) had 5, and *C. tropicalis* (lowest variation) had 4 (**Figure 3A**), corresponding to 19, 31, and 25 isotype-switched tRNA genes, respectively.

To assess whether fixed isotype switches reflect ancient, likely neutral or beneficial events, we generated secondary structure-aware multiple sequence alignments of all alleles per amino acid backbone across species (Methods). These alignments enabled visualization of allele sequence relationships. For example, leucine-backbone genes switched to isoleucine in all strains of all three species, and these isoleucine-encoding alleles cluster together across species rather than with leucine-retained alleles from their respective species (**Figure 3B**). Similarly, another set of leucine-backbone genes' alleles switch to glycine in all three species and clustered together across species (**Figure 3B**). These examples support the hypothesis of fixed isotype switches deriving from ancient, shared events. Interestingly, at one leucine backbone gene, *C. elegans* I.trna17, 80.2% of the strains harbored alleles with a glutamine-recognizing anticodon (TTG in the DNA) while the remain 19.8% of strains had anticodons matching isoleucine (TAT): an anticodon switch away from leucine was fixed at this gene, but different strains in the population have tRNAs matching different amino acids at this gene.

Next, we investigated the 'some switch, some not' genes to investigate intra-specific variation in isotype switching. These switches tended to be species-specific and low frequency, possibly reflecting deleterious new mutations that may soon be purged (**Figure 3A, Table S5**). Across species, the extent of low-frequency isotype switching generally mirrored overall genetic and tRNA allelic variation: *C. briggsae* showed the most and *C. tropicalis* showed the least (**Figure**

3A). This observation supports the idea that isotype switching may arise from the high mutation load imposed on tRNAs by TAM; indeed, anticodons frequently harbor genetic variation (**Figure 1C**).

To better understand contemporaneous isotype switching in individual tRNA genes, we examined all methionine alleles (**Figure 3C**). Methionine switches to threonine in a singleton (observed in one strain) allele in *C. elegans* and to isoleucine in a singleton allele in *C. tropicalis* (**Figure 3C**). These rare alleles did not cluster closely on the allele tree, suggesting they arose as independent, more recent events (**Figure 3C**).

We observed only one case in which an isotype switch was neither rare nor fixed, *C. elegans* tRNA gene V.trna22 (**Table S5**). This predicted valine (backbone) gene had an anticodon corresponding to methionine (CAT) in 546 strains (79.7% of the population), including the reference strain. In contrast, a minority of 118 strains (12.3%) harbored alleles at this gene encoding the backbone-matching valine anticodon (CAC). This example highlights that *Caenorhabditis* species form fertile ground to study the full spectrum of isotype switching, from ancient events to ongoing variation.

Distinctive genomic organization of tRNA genes

Genomic context can provide important clues about evolutionary processes, so examined how tRNA genes were organized relative to each other and to genome architecture in *Caenorhabditis* genomes. Here, we used the positions of all reference genome tRNAs, focusing on this one representative set per species rather than within-species variation.

We first asked whether tRNA gene distribution followed the same location rules as protein-coding genes in in *Caenorhabditis* genomes. These worms' holocentric chromosomes exhibit distinct recombination domains with frequent recombination in the chromosome arms and less frequent recombination in chromosome centers (Barnes et al. 1995; Hillier et al. 2007; Rockman and Kruglyak 2009; Ross et al. 2011; Kaur and Rockman 2014; Bernstein and Rockman 2016; Noble et al. 2021; Stevens et al. 2022). Commensurately, gene density is generally higher and genetic variation lower in the more stable center chromosomal regions than in chromosome arms (Barnes et al. 1995; Hillier et al. 2007; Rockman and Kruglyak 2009; Noble et al. 2021), a pattern visible by eye for protein coding genes (**Figure 4A**). Surprisingly, tRNA genes did not follow this rule: they were distributed more evenly across chromosomes, sometimes even enriched in chromosome arms (**Figure 4A**). This tRNA-specific gene distribution across recombination domains was evident (and significantly different from that of protein-coding genes) on each of the six chromosomes (five autosomes and X) in each species (**Figure 4A**).

In all three species, tRNA genes occurred on all chromosomes in all center and arm recombination domains (**Figure 4, Figure S5**). All isotypes comprising more than one tRNA gene had genes on at least half of all chromosomes (three of six) (**Figure 4, Figure S5**). This genome-wide distribution raised the question of whether tRNA genes were randomly distributed in the genome regardless of their relatedness to each other, or if more similar genes (at the isotype, isodecoder, or full sequence level) were non-randomly located. In favor of randomness, genes with identical sequence occur on different chromosomes. For example, in each species, 9 genes found on three (*C. tropicalis*) or four (*C. elegans, C. briggsae*) chromosomes coded for

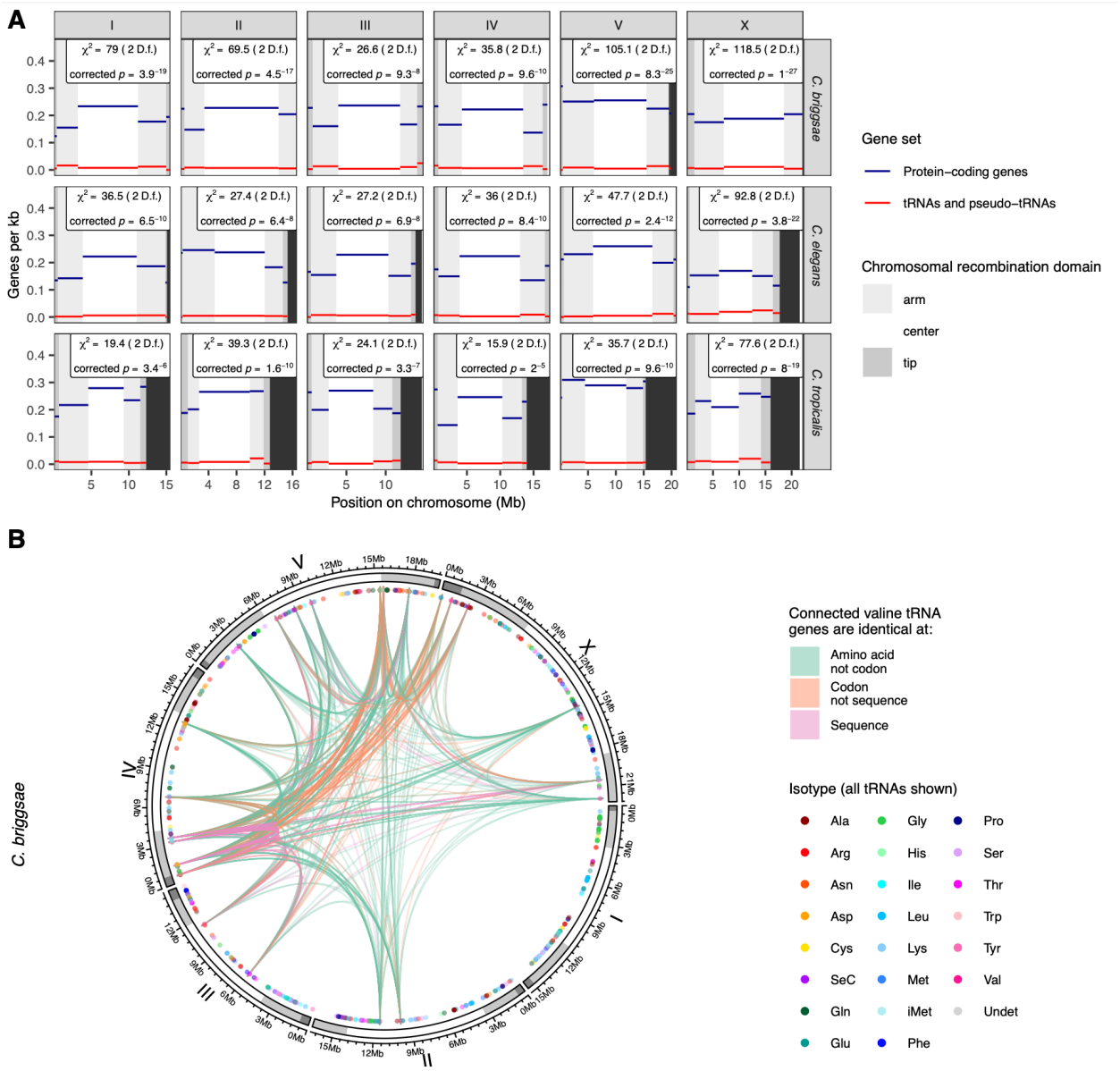


Figure 4. Distribution of tRNA genes within *Caenorhabditis* genomes. **A,** Per-kb density of protein-coding and tRNA genes in each chromosome's high recombination (arms) and low recombination (center) domains. Species' chromosomes have different lengths; where one species has a longer chromosome than another, the other species' plotting region is blacked out. Results from per-species, per-chromosome χ^2 tests comparing protein-coding and tRNA gene distribution across the domains are annotated on the plot; p-values are Bonferroni corrected for 18 (chromosomes x species) tests. **B,** The distribution of all tRNA genes in the *C. briggsae* reference genome (colored dots, n = 779 non-all-nonfunctional allele, single predicted amino acid backbone tRNAs) shown with pairwise connections between all valine tRNA genes (colored lines, n connections = 595 between n = 35 valine tRNA genes). Chromosomes and genomic coordinates are shown in circular layout; chromosome shading corresponds to chromosomal recombination domain (see **A**).

initiation methionine (iMet). Despite their wide genomic distribution, in *C. elegans* and *C. tropicalis*, all 9 iMet tRNA genes had identical sequences in the reference genome, while in *C.*

briggsae, two distinct sequences corresponded to iMet. Other isotypes such as valine (**Figure 4B, Figure S5**) contained multiple sequence classes, with both highly similar genes (identical sequences) and more divergent genes (different isodecoders) residing on different chromosomes from one another.

Nevertheless, in all three species, tRNA gene pairs were more likely to be on the same chromosome the more similar they were: genes within an isotype were more likely to be on the same chromosome than genes with different isotypes; genes within an isodecoder class (same anticodon) were more likely to share a chromosome than genes within the same isotype but with different anticodons; and genes with full-length identical sequence were more likely to share a chromosome than genes in the same isodecoder class that had other (non-anticodon) sequence differences (**Figure S6**). Furthermore, when on the same chromosome, tRNA gene pairs tended to be closer together the more similar they were, following the same hierarchy and also when examining pairwise sequence divergence, with sequence similarity predicting genomic proximity (**Figure S7**).

Discussion

tRNA pools exhibit remarkable macro-similarity across eukaryotes, even as they experience immense evolutionary churn in gene number, sequence, and isotype identity (Rogers et al. 2010; Wang and Ruvinsky 2012; Rogers and Griffiths-Jones 2014; Velandia-Huerto et al. 2016; Yona et al. 2013). Here, we show that the tRNA pools of three *Caenorhabditis* species, representative of the tRNA pools found in eukaryotes generally, harbor extensive intra-specific variation in tRNA genes. We propose that this pattern likely contributes to tRNA functional diversity, evolutionary fitness, and mutation load in *Caenorhabditis* and across eukaryotes.

Perspective

Our nascent understanding of natural genetic variation in tRNA repertoires, which this work aims to advance, points to a broad, biologically significant field of study with untested predictions. Perhaps most importantly, the field needs to ascertain how and whether different tRNA repertoires within a species confer fitness differences among individuals. Single-tRNA mitochondrial and selenocysteine nuclear examples demonstrate that certain tRNA mutations can have deleterious effects (Florentz et al. 2003; Yarham et al. 2010; Suzuki et al. 2011; Schoenmakers et al. 2016; Lant et al. 2019). However, analysis of nuclear encoded tRNAs adds significant analytical and biological complexity given high copy number and multidimensional variation. Work connecting specific molecular effects of tRNA modifications (such as (Torres et al. 2019; Davey-Young et al. 2024; Tennakoon et al. 2025)) to larger-scale organismal phenotypes in experimentally tractable systems across multiple generations will be particularly valuable.

To this end, an intermediate key question is which specific tRNAs have outsized use and functional impact in a given species or individual and which are interchangeable or redundant. In mammals (Torres 2019; Hughes et al. 2023; Gao et al. 2024) and yeast (Bloom-Ackermann et al. 2014), recent work suggests that specific genes of isoacceptor and isodecoder classes exert disproportionate functional impact, despite evidence of functional redundancy within isoacceptor classes (Leung et al. 1991; Kutter et al. 2011). Understanding these patterns—and, critically, how variation perturbs them—will be important. Integrative computational models that combine information about tRNA gene chromatin conformation and predicted and empirical expression with canonical tRNA function models (*e.g.* (Thornlow et al. 2020)), and variation atlases like the one developed here, could generate predictions of context-specific tRNA gene function for empirical testing. Because tRNA gene variants may themselves alter expression and activity, direct quantification and epigenetic profiling (that distinguishes identical genes via flanking sequence awareness), in individuals with diverse allelic repertoires, ideally across multiple tissues and life stages, will be critical.

A related avenue of inquiry is to intersect tRNA repertoires with codon usage bias. Certain tRNA allelic combinations might be particularly disfavored in the context of species-specific codon usage overall or within specific gene sets. Codon usage itself might also vary within a species. Whether codon usage evolves or varies within species, and whether and how tRNA repertoire differences cause or result from such variation, will be important questions to address.

Comments on this study

A central phenomenon in tRNA evolution is transcription-associated mutagenesis (TAM), which underlies mutation patterns at tRNA genes and their flanking regions (Thornlow et al. 2018). We recently demonstrated, using a novel sequence substitution model, that standing variation in *C. elegans* nuclear tRNAs is well explained by mutational input from TAM (Banos et al. 2025). Here, we show evidence of a strong TAM signature at flanking regions of tRNA loci in *C. elegans* strains as well as two more nematode species, *C. tropicalis* and *C. briggsae* (**Figure 1C**). These findings replicate initial reports in other eukaryotes and further emphasize the universality of this phenomenon (Thornlow et al. 2018). Because tRNA genes are transcribed in all cells, TAM presumably acts organism-wide, consistent with the high somatic mutation burden in tRNA genes in human cancer and aging (Murillo-Recio et al. 2025). However, only germline TAM generates the heritable variation shaping tRNAs. This reasoning suggests that TAM is universal despite diverse mechanisms of eukaryotic germline formation, reinforcing predictions that population-level variation in tRNA genes, like we observe, imposes a mutation load on genomes across taxa.

Within tRNA gene bodies, we do not detect obvious variation in strength of purifying selection; we observe mutations in each tRNA secondary structure component. One interpretation is that all tRNA mutations are similarly deleterious; a more plausible explanation is that all sites experience high mutation rates and overall selection is stronger than any site-specific differences we can currently detect. Additional samples and data may enable finer-scale inference of selection differences, as recently achieved for somatic tRNA mutations in human aging and cancer (Murillo-Recio et al. 2025) and via saturating mutagenesis and experimental evolution in yeast (Li et al. 2016). Furthermore, selfing in nature in these species may slow the purging of deleterious (Loewe and Cutter 2008), increasing their detectability as standing variation while potentially dampening signals of selection.

Much of the tRNA allelic variation we report here likely impacts tRNA function by affecting folding and structure or amino acid proofreading or identity (**Figure 2**). Our analysis indicates that strains differ in their number of functional tRNA genes. Because we analyzed only tRNA genes present in the reference genome and did not assess copy-number variation at these loci or search for novel tRNAs with distinct sequences or genomic locations in non-reference strains, we cannot explicitly define the number of predicted functional tRNA genes for each strain. Given that copy-number variation of tRNA genes is common, its analysis will be helpful for future work (Iben and Maraia 2012, 2014; Darrow and Chadwick 2014; Velandia-Huerto et al. 2016; Hughes et al. 2023). To fully resolve the nuclear genome tRNA complement of each strain would require a copy number and allele-aware analysis, perhaps based on long-read, strain-specific assemblies (*e.g.* of the Hawaiian strain CB486, (Thompson et al. 2015; Kim et al. 2019)). Such a study must carefully distinguish true genomic differences in tRNA complement from assembly-induced differences that can arise even between assemblies of the same strain or organism (as recently demonstrated in humans (Murillo-Recio et al. 2025)). An additional methodological improvement would be to map gene identity (orthology) across species, which would enable finer-grained evolutionary analysis. Because tRNA genes are short, multi-copy, and fast evolving, inferring orthology is not straightforward. A leading option for such an analysis

uses synteny to define orthologs. Given that the three species analyzed here are separated by millions of years of evolution, examining more closely related 'sister species' (*e.g.*, *C. briggsae* and *C. nigoni* or *C. latens* and *C. remanei*) may be more tractable for synteny analysis of highly variable, fast evolving tRNA genes. A smaller scale analysis option would be to examine the ~43% of *C. elegans* genes contained in other genes' introns by matching host genes to their orthologs in the other species (Wang and Ruvinsky 2012)

Isotype switching is particularly interesting in considering changes to tRNA function. We observed two broad categories of isotype switching (**Figure 3**). First, we detected population-fixed isotype switches that likely predate within-species and likely sample the isotype remolding reported across eukaryotes (Rogers and Griffiths-Jones 2014; Velandia-Huerto et al. 2016). Leucine-backbone tRNAs most often carried switches, with many genes fixed for other anticodons in all three *Caenorhabditis* species. This pattern suggests that these switches predate speciation and perhaps fixed neutrally or adaptively, consistent with observations of leucine observation in nematodes, humans, and other taxa (Lant et al. 2019; Rogers and Griffiths-Jones 2014), but confirming this interpretation would require resolving orthology and synteny. Because leucine's aminoacyl-tRNA synthetase (aaRS), like those for alanine and serine, does not rely on the anticodon for amino acid identity, leucine tRNAs can and do deliver leucine even with anticodons for other amino acids (Giege et al. 1998; Davey-Young et al. 2024). Fixed isotype switching also occurred from alanine and serine backbones, consistent with further possible species-wide isotype remolding. Second, we detected rare, polymorphic isotype switches present in one or a few strains in all three species. Based on their tRNA backbones, these observed switches likely do not cause widespread amino acid mischarging or misincorporation, but could slow amino acid loading and increase energy-intensive proofreading by aminoacyl-tRNA synthetases (Perona and Gruic-Sovulj 2014). Taken together, we predict that the tRNA complements of some strains are suboptimal due to isotype switching; some strains are burdened by each isotype switch, others are not. Determining the functional consequences of these polymorphic switches will clarify the contributions of tRNA gene genetic variation to organismal fitness and mutational load.

While broad scale tRNA complement and variation patterns were shared across species and likely generalize across eukaryota, *C. briggsae* exhibited notably more tRNA gene variation than the other species. This excess exceeds that which can be explained by its slightly larger sample size and is present across the genome (*i.e.*, not limited to tRNAs). *C. briggsae* also has the highest effective population size (N_e) of the three species in this study. Higher N_e should elevate neutral diversity; conversely, higher N_e should also strengthen purifying selection and reduce deleterious variation (Charlesworth 2009); in any event, the relationship between N_e and the nucleotide variation observe at tRNAs is not immediately clear. While we cannot rule out technical explanations for the apparent excess of variation in *C. briggsae*, they seem unlikely: all three species' genomic data was processed with the same high-quality pipelines (Crombie et al. 2024) and hyperdivergent regions that complicate variant calling characterize both *C. briggsae* and *C. elegans* (Thompson et al. 2015; Lee et al. 2021; Stevens et al. 2022; Moya et al. 2025).

The standing variation at tRNA genes that we report here may persist longer in these three nematode species because of their androdioecious mating system; selfing limits recombination

to roughly one crossover per chromosome per generation, often between near-identical genotypes (Loewe and Cutter 2008). How quickly the deleterious variants at tRNA are purged remains unclear; background selection is a dominant force in these genomes, and varies between the low-recombination centers of chromosomes and the higher-recombination arms (Rockman et al. 2010; Andersen et al. 2012), where tRNAs are disproportionately distributed (**Figure 4**). We hypothesize that this distribution also reflects rapid birth-death cycles that characterize tRNA gene families (Nei and Rooney 2005; Wang and Ruvinsky 2012) as well as their duplication mechanisms. tRNAs likely duplicate as mobile genetic elements; indeed, many SINE elements originate from tRNA genes. Different mobile element-derived sequence families have different distributions in the *C. elegans* genome, with some center-biased and others are arm-biased (Surzycki and Belknap 2000; Laricchia et al. 2017), so the tRNA-specific distribution might reflect tRNA-specific evolution.

The tRNA gene variation reported here may well confer biological significance beyond the canonical role of tRNA in translation; alternative activities of tRNAs continue to be discovered (reviewed in (Schimmel 2018)). Furthermore, because tRNAs contain internal promoters, tRNA gene variation might change when or where a tRNA is expressed (Galli et al. 1981; Giege et al. 1998). Variation may also affect the post-translational tRNA modifications that play a central role in tRNA regulation and function (Schimmel 2018). Beyond translation, tRNA gene sequence differences could influence tRNA participation in protein complexes and the generation of tRNA fragments. Indeed, human isodecoder gene set variation has been observed to impact expression levels of tRNA fragments, which have diverse roles, more than full-length tRNAs (Torres 2019; Torres et al. 2019)**.**

Conclusion

Here, we uncovered remarkable genetic variation in nuclear-encoded tRNA genes in three *Caenorhabditis* species, much of which may affect tRNA function. The insights, datasets, and resources presented here, including an interactive web app available at https://wildworm.biosci.gatech.edu/trnarefalleles/, will serve as a springboard for investigation of the causes and consequences of nuclear tRNA variation across eukaryotes.

Materials and Methods

The code written for this study is available at https://github.com/paabylab/trnavcfalleles.

Genomic data sources

The reference genome for *C. elegans* was version ws283 from WormBase (Sternberg et al. 2024), the same as used for short-read sequencing data alignment and variant calling by the *Caenorhabditis* natural diversity resource (CaeNDR, (Crombie et al. 2024)) in the *C. elegans* 687 sample (strain) VCF used here (20250625 release, https://caendr.org/data/data-release/c-elegans/20250625). For *C. briggsae*, reference genome fasta and 719 sample (strain) VCF were downloaded from CaeNDR (20250626 release, https://caendr.org/data/data-release/c-briggsae/20250626; reference genome build from (Stevens et al. 2022)). Likewise, for *C. tropicalis*, we obtained the reference genome fasta and XXX sample (strain) VCF from CaeNDR (20250627 release, https://caendr.org/data/data-release/c-tropicalis/20250627; reference genome build from (Noble et al. 2021)). Specifically, we used the hard-filtered, isotype-only VCFs in all cases; these VCFs include only wild strains that are not genetically identical and only stringently filtered variants. We also obtained used independent genome sequence from one additional (non-reference) *C. elegans* strain CB4856, generated *de novo* from long-read sequencing ((Thompson et al. 2015), NCBI GenBank JZEW00000000).

Reference genome tRNA identification

To identify tRNA genes and tRNA-like sequences, we ran tRNAscan-SE 2.0 (v2.0.12) (Chan and Lowe 2019; Chan et al. 2021) on reference genomes for the three *Caenorhabditis* species as well as the CB4856 strain genome. We ran tRNAscan-SE using default settings, including full logs, progress reports, and isotype-specific modeling by including options *--log, --progress, --detail*, and *--isospecific*. We used the tRNAscan-SE output for all downstream analysis: it gave us the codon sequence; the cognate amino acid for this codon; the predicted cognate amino acid based on evolutionary analysis of the tRNA backbone; the Infernal (Nawrocki and Eddy 2013) predicted tRNA folding score; a score of goodness-of-fit to the isotype (amino acid) assigned by the codon (related to the predicted backbone); and a note including if tRNAscan-SE predicted the reference sequence to be a pseudogene. We used these outputs rather than data available on gtRNAdb (Chan and Lowe 2016) in order to use the latest reference genomes available and ensure consistency in our data processing and interpretation across genomes and species. Reference-sequence genes were predicted to be functional tRNAs when they were not called pseudogenes by tRNAscan-SE, had predicted codons, and had Infernal scores > 20 (usually much higher); in downstream analyses we interrogated all alleles at the gene for functional status and therefore could also filter tRNA genes by if all their alleles were predicted to be nonfunctional. This analysis errs on the side of including sequences as possible tRNAs vs. excluding them or only analyzing very filtered sets of genes given our interest in mutational variance in long and short evolutionary timescales at tRNAs.

Determining variation in tRNA flanking regions

We first subset each species's whole-genome VCF to include only the variants within 1 kb of any tRNA gene or tRNA-like sequence identified above using bcftools (v1.11-19) (Danecek et al.

2021), then determined recorded the identity of each variant that occurred within 40bp up or downstream of each tRNA gene; we also tracked the strains in which this variant was observed or had missing genotype. Both SNVs and INDELs were included. (Note that all three species are androdiecious and self-mate in nature, so the wild isotypes are naturally inbred homozygous genome-wide; for this reason, we counted any variant called as heterozygous as missing a genotype in that strain.) From these data we computed the proportion of tRNAs with an observed variant at each flanking basepair from -40 upstream of the gene to 40 downstream of the gene (appropriately accounting for strandedness), overall and restricting to certain gene sets and mutation types.

Code relevant for these analyses is available in our github repository in directory *flank_specific*.

Allele frequency analyses of variants tRNA and protein-coding exon regions

To characterize SNV and INDEL variants in tRNA genes and protein-coding exon regions, we extracted all biallelic variants in these regions from the CaeNDR VCFs. We note that while tRNA genes are highly similar, they are short, so the short-read 90+ bp paired-end Illumina sequencing data underpinning the variant calls will include both tRNA genes and their more unique genomic contexts in the majority of cases. tRNA gene regions were all those in the .bed files generated by tRNAscan-SE (tRNA and tRNA-like genes), and we generated protein-coding exonic regions .bed files from the species' GTFs. We used bcftools bcftools (v1.11-19) (Danecek et al. 2021) to extract the counts of strains with homozygous reference, homozygous alternate, heterozygous (rare in these isogenic species), and missing genotypes for each biallelic variant in a region of interest, retaining the region's identity. We then used these genotype counts to investigate minor allele frequency (MAF) distributions between tRNA and protein-coding regions; these comparisons' results were robust to calculation of MAF as the number of minor allele-homozygous genotypes divided by the total number of strains or divided by the number of strains with non-missing genotypes at that particular variant. Tajima's *D* (Tajima 1989) was calculated from the site frequency spectrum generated from these genotype counts (following (Yang 2022)). We performed this calculation restricting to sites with >80% or >90% samples with non-missing genotypes as missingness impacts the calculation and based only on SNVs or also including INDELs; results were robust to all variations.

Code relevant for these analyses is available in our github repository in directory *variant_characterization.*

Gene location analyses

Analyses of gene location with regard to recombination domain used recombination domain genomic boundaries for *C. elegans* from (Rockman and Kruglyak 2009) (Table 1), for *C. briggsae* from (Stevens et al. 2022) (Table 2), and for *C. tropicalis* from (Noble et al. 2021) ('ctbp' data in https://github.com/lukemn/tropicalis/blob/master/geneticMap/data/c_tropicalis_flye_geneticMap.rda). All protein-coding genes for each species for comparison were obtained from the GTF files matching the reference genomes (ws283 from WormBase narrowed to the protein coding biotype for *C. elegans*; protein-coding only gene models generated by CaeNDR for *C. briggsae* and *C. tropicalis,* canonical_geneset.gtf files). We assigned genes to specific recombination

domains based on the domain that encompassed the midpoint of the gene. From these assignments, we computed the per-kb gene density of each class of gene in each recombination region; the difference in domain distribution between genes within species and chromosomes (*i.e.*, arm vs center gene density); and the difference in domain distribution between the gene sets within species and chromosomes (*i.e.*, differential arm vs center gene density between two species).

To examine the pairwise location distribution of tRNA genes, we used the gene's midpoint as its single location value and, for each pair of tRNA genes in the reference genome, determined whether the pair was on the same chromosome and if so, how far apart the midpoints were. Then each pair was classified in various ways relevant to its analysis (whether the gene pair had identical sequence; whether the gene pair was from the same isotype; the sequence alignment divergence between the gene pair [see below]; etc).

Code relevant for these analyses is available in our github repository: *analysis/trna_location_analyses.R.*

tRNA allele identification and characterization

To identify all tRNA alleles in the three populations, we narrowed the VCFs to include only tRNA genes and surrounding regions (identified by tRNAscan-SE as described above), then generated FASTA files containing the sequence of each tRNA allele observed by inserting each strain's observed variant(s) into the appropriate position in the tRNA gene region, overwriting the reference base at that position (SNVs and INDELs included). We kept track of which strains observed which variants and harbored which alleles, and when strains had missing genotypes at any of the variants in a given tRNA we did not assign that strain an allele identity for that tRNA. To characterize all of these alleles, we re-ran tRNAscan-SE on the FASTA file containing all tRNA allele sequences and used this output plus knowledge of the input alleles that were not identified as possible tRNAs by tRNAscan-SE (options as described in the section discussing running tRNAscan-SE on the reference genomes). Therefore, each allele was annotated with its predicted isotype (codon based) and amino acid backbone, as well as all other tRNAscan-SE output information. Each allele was given predicted functional classifications based on the following criteria:

1) 'Best': an allele was classified as 'best' if its isotype matched its predicted amino acid backbone (identical 'AA' and 'AlleleCM' columns in tRNAscan-SE output) and it had the best (highest and putatively functional) infernal score of any allele at this tRNA (and was not called a pseudogene by tRNAscan-SE or missed in tRNAscan-SE's calling)
2) 'Functional': as 'Best', but with a lower infernal score (still >>20)
3) 'Isotype switch': an allele was classified as 'isotype switch' if it was presumed functional (not meeting any 'nonfunctional' criteria, below) and its isotype did not match its predicted amino acid backbone ('AA' and 'AlleleCM' columns in tRNAscan-SE output differed)
4) 'Nonfunctional': an allele was classified as 'nonfunctional' if it was called a pseudogene by tRNAscan-SE, missed in tRNAscan-SE's calling, had an indecipherable/NNN codon,

or had an infernal score < 20 (in practice all with such scores were flagged as pseudogenes).

These calls were used in further gene-level analyses, such as identifying genes with all pseudogenized/nonfunctional alleles and classifying genes as having fixed isotype switching or variable isotype switching.

To estimate confidence in our isotype switching classifications specifically, we intersected observed genetic variation (from the VCF) with tRNAscan-SE-derived allele annotations. At genes where all alleles were called isotype switches, no genotype variability exists, so could not interpret how often we could pinpoint genotype changes 'used' by tRNAscan-SE. Therefore, we focused on genes with segregating alleles that differ in their cognate amino acid/isotype switch identity. Among these, 56-61% of variable (low-frequency) isotype switches were attributed to anticodon mutations and 39-44% to backbone mutations. In all backbone mutation cases and 58-80% of the anticodon isotype switch cases, we could clearly map variants in the VCF to the predicted isotype switch. Some of the unexplained isotype switch calls likely stem from insertion-deletion polymorphisms, which we included in tRNA allele generation but excluded from variant analysis. Combined with the fact that most high-frequency isotype switches were anticodon based and necessarily excluded from this analysis, we infer that these numbers represent a conservative lower bound on the reliability of isotype switch identification.

Code relevant for these analyses is available in our github repository: see especially all scripts in *generate_alleles*; *analysis/show_mutational_variation.R; analysis/charisotypeswitching.R.*

Secondary-structure aware tRNA sequence alignment and related analyses

To enable secondary structure-aware alignment of tRNA gene sequence, as these structures do not have invariant lengths across sequences (Chan and Lowe 2016), we used the secondary structure predictions from tRNAscan-SE to determine for each tRNA gene in the reference genome and each allele which base mapped to which structure. We first excluded introns from this analysis, excising the tRNAscan-SE predicted intron positions from the tRNA gene sequence. For example, we first identified the codon, as tRNAscan-SE gives its position in its secondary structure output, then inferred the surrounding unpaired regions were part of the anticodon loop, the paired regions surrounding them were part of the anticodon stem region, etc. In some cases, especially pseudogene, we could not map all bases in the tRNA to a predicted secondary structure, so we flagged this and excluded them from downstream analyses. In other cases, predictions were imperfect, for example not all bases in the acceptor stem were paired; since this can and does happen in biology, we retained these in analyses while flagging them. We determined that including vs. excluding all genes with such tags did not yield obviously different results, so we included them in the results provided here.

We used the same logic to identify the relative position within secondary structure of each VCF variant: we broke the reference tRNA up into its secondary structure components, mapped genomic positions to the position of each component, and recorded the relative position in the relavent secondary structure. This enabled us to generate estimates of the per-structure proportion of tRNAs with variants and normalize this to the average length of the structure across all tRNAs.

We generated multiple sequence alignments of all alleles observed across all three species populations in a cloverleaf-structure aware fashion by breaking each sequence into its individual structural components as specified above, then aligning each structure separately, then finally concatenating the per-structure alignments into full tRNA alignments. Specifically, we ran *foursale* (4SALE CMD version, v1.3) (Wolf et al. 2014), which uses clustalw2 (v2.1) (Larkin et al. 2007) to align sequences after combining the RNA secondary structure information and base information at each base. Parameters were *-gapextend 1* and *-gapopen 1*. This was run separately for each consecutive region of the tRNA: the acceptor stem's left (first) paired part to the beginning of the D arm; the D arm (stem and loop) to the beginning of the anticodon arm; the anticodon arm with the anticodon N-masked to avoid anticodon misalignment; separately the anticodon alone (to force anticodon alignment); the variable arm; the T arm up to the acceptor stem; and then the remaining part of the acceptor stem and the acceptor stem overhang. These separate alignments were concatenated in this order, with the anticodon-specific alignment replacing the N-masked anticodon in the anticodon arm alignment, to get the full-length, all-allele tRNA multiple sequence alignment.

These alignments were used 1) for determining the sequence alignment distance for all pairs of unique sequence and 2) for generating allele trees. The sequence alignment distance or sequence divergence was calculated in two ways, only for unique sequences of alleles in the reference genome: first, simply the number of character differences between two alignments (so a gap is counted as different from any base, and different bases are counted as different from each other); second, the number of distinct stretches of difference between two alignments. Neighbor-joining trees were generated from all alleles or subsets of alleles (from within a species; from a specific amino acid background; etc) using maximum-likelihood sequence alignment distances (functions *dist.ml* and *NJ* from the *phangorn* R package (v2.12.1) (Schliep 2011). Trees were further annotated, manipulated, and plotted using the *ape* (v5.8-1) (Paradis and Schliep 2019), *ggtree* (v3.16.3) (Yu et al. 2017), and *tidytree* (v0.4.6)(Yu 2022) R packages.

General software tools used for analyses and figures

Tools used for specific analytical purposes are often described in the relevant sections; here, we share tools used for general thinking, data processing, and figure creation.

We used Georgia Tech's Enterprise access to Microsoft Copilot (running GPT 4 and 5) for occasional problem solving and text editing suggestions (not for any substantial idea or text generation).

Analysis scripts were written in R (up to v4.5.1) (R Core Team 2025) and Python (v3.7) (www.python.org). Workflow scripts were written and run using Nextflow (v22.10.7) (www.nextflow.io). Compute-intensive analyses and workflows were run via the Partnership for an Advanced Computing Environment (PACE), the high-performance computing environment at the Georgia Institute of Technology.

Data wrangling included R packages *data.table* (v1.17.8) (Dowle and Srinivasan 2022), *argparser* (v0.7.2) (Shih 2021), *formattable* (v0.2.1) (Ren and Russell 2021), and *RRNA* (v1.2) (Bida and Maher 2012) and Python packages included pyvcf (v0.6.4) and numpy (v1.23.0) (Harris et al. 2020). Data display and figure creation R packages included *ggplot2* (v3.5.2)

(Wickham 2016), *cowplot* (v1.2.0) (Wilke 2020), *RColorBrewer* (v1.1-3) (Neuwirth 2022), *ggforce* (v0.4.0) (Pedersen 2022), *ggsignif* (v0.6.4) (Ahlmann-Eltze and Patil 2021), *ggpattern* (v1.1.4) (FC et al. 2022), *ggpmisc* (v0.6.2) (Aphalo 2024), *ggnewscale* (0.4.2) (Campitelli 2025), *ggh4x* (v0.3.1) (van den Brand 2025), *gridBase* (v0.4.7) (Murrell 2014), *scales* (v1.4.0) (Wickham et al. 2025), *scatterpie* (v0.2.5) (Yu 2025), *circlize* (v0.4.16) (Gu et al. 2014), *magick* (v2.8.7) (Ooms 2025a), and *pdftools* (v3.5.0) (Ooms 2025b).

Data Availability

Datasets generated in this study are available as supplemental material or via the Zenodo repository at https://doi.org/10.5281/zenodo.17237304. Data is interactively available via web app at https://wildworm.biosci.gatech.edu/trnarefalleles/. Code used in this study is available at https://github.com/paabylab/trnavcfalleles.

Acknowledgments

The authors thank Abigail Lind and members of the Lind lab, Rohini Janivara, Abbe LaBella, and members of the Southeast Center for Mathematics and Biology for helpful conversations about this work. The Georgia Tech College of Sciences' Academic and Research Computing Services team provided web server configuration support for the interactive web app. Research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment at Georgia Tech supported this research. This work was made possible by publicly available data from CaeNDR (https://caendr.org/) and the many members of the research community who have collected and shared strains.

Funding

This work was funded by National Science Foundation EDGE grant 2319796 (to ABP), by the NSF-Simons Southeast Center for Mathematics and Biology (SCMB) through the grants National Science Foundation (NSF) DMS-1764406 and Simons Foundation/SFARI 594594, and by support from the Georgia Institute of Technology.

References

Ahlmann-Eltze, C., and I. Patil, 2021 ggsignif: R Package for Displaying Significance Brackets for 'ggplot2'. *PsyArxiv*. https://psyarxiv.com/7awm6. 10.31234/osf.io/7awm6.

Andersen, E.C., J.P. Gerke, J.A. Shapiro, J.R. Crissman, R. Ghosh *et al.*, 2012 Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nat Genet* 44 (3):285–290. https://www.ncbi.nlm.nih.gov/pubmed/22286215. 10.1038/ng.1050.

Andersen, E.C., and M.V. Rockman, 2022 Natural genetic variation as a tool for discovery in Caenorhabditis nematodes. *Genetics* 220 (1). https://www.ncbi.nlm.nih.gov/pubmed/35134197. 10.1093/genetics/iyab156.

Aphalo, P.J., 2024 ggpmisc: Miscellaneous Extensions to 'ggplot2'. https://CRAN.R-project.org/package=ggpmisc. https://CRAN.R-project.org/package=ggpmisc.

Banos, H., L. Wang, C. Simonti, A. Paaby, and C. Heitsch, 2025 How much is Transcription-associated Mutagenesis Driving tRNA Microevolution? *bioRxiv*:2025.2003.2026.645593. https://www.biorxiv.org/content/biorxiv/early/2025/03/31/2025.03.26.645593.full.pdf. 10.1101/2025.03.26.645593.

Barnes, T.M., Y. Kohara, A. Coulson, and S. Hekimi, 1995 Meiotic recombination, noncoding DNA and genomic organization in Caenorhabditis elegans. *Genetics* 141 (1):159–179. https://www.ncbi.nlm.nih.gov/pubmed/8536965. https://www.ncbi.nlm.nih.gov/pubmed/8536965.

Barriere, A., and M.A. Felix, 2005a High local genetic diversity and low outcrossing rate in Caenorhabditis elegans natural populations. *Curr Biol* 15 (13):1176–1184. https://www.ncbi.nlm.nih.gov/pubmed/16005289. 10.1016/j.cub.2005.06.022.

Barriere, A., and M.A. Felix, 2005b Natural variation and population genetics of Caenorhabditis elegans in *WormBook*. http://wormbook.org/chapters/www_naturalvariationgenetics/naturalvariationgenetics.html. 10.1895/wormbook.1.43.1.

Berg, M.D., D.J. Giguere, J.S. Dron, J.T. Lant, J. Genereaux *et al.*, 2019 Targeted sequencing reveals expanded genetic diversity of human transfer RNAs. *RNA Biol* 16 (11):1574–1585. https://www.ncbi.nlm.nih.gov/pubmed/31407949. 10.1080/15476286.2019.1646079.

Bermudez-Santana, C., C.S.-O. Attolini, T. Kirsten, J. Engelhardt, S.J. Prohaska *et al.*, 2010 Genomic organization of eukaryotic tRNAs. *BMC Genomics* 11 (1):270. https://doi.org/10.1186/1471-2164-11-270. 10.1186/1471-2164-11-270.

Bernstein, M.R., and M.V. Rockman, 2016 Fine-Scale Crossover Rate Variation on the Caenorhabditis elegans X Chromosome. *G3 (Bethesda, Md.)* 6 (6):1767–1776. https://www.ncbi.nlm.nih.gov/pubmed/27172189. 10.1534/g3.116.028001.

Bida, J.P., and L.J. Maher, 3rd, 2012 Improved prediction of RNA tertiary structure with insights into native state dynamics. *RNA* 18 (3):385–393. https://www.ncbi.nlm.nih.gov/pubmed/22279150. 10.1261/rna.027201.111.

Bloom-Ackermann, Z., S. Navon, H. Gingold, R. Towers, Y. Pilpel *et al.*, 2014 A comprehensive tRNA deletion library unravels the genetic architecture of the tRNA pool. *PLoS Genet* 10 (1):e1004084. https://www.ncbi.nlm.nih.gov/pubmed/24453985. 10.1371/journal.pgen.1004084.

Campitelli, E., 2025 ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'. https://CRAN.R-project.org/package=ggnewscale>. https://CRAN.R-project.org/package=ggnewscale>.

Chan, P.P., B.Y. Lin, A.J. Mak, and T.M. Lowe, 2021 tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res* 49 (16):9077–9096. https://www.ncbi.nlm.nih.gov/pubmed/34417604. 10.1093/nar/gkab688.

Chan, P.P., and T.M. Lowe, 2009 GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 37 (Database issue):D93–97. https://www.ncbi.nlm.nih.gov/pubmed/18984615. 10.1093/nar/gkn787.

Chan, P.P., and T.M. Lowe, 2016 GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* 44 (D1):D184–189. https://www.ncbi.nlm.nih.gov/pubmed/26673694. 10.1093/nar/gkv1309.
Chan, P.P., and T.M. Lowe, 2019 tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol* 1962:1–14. https://www.ncbi.nlm.nih.gov/pubmed/31020551. 10.1007/978-1-4939-9173-0_1.
Charlesworth, B., 2009 Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10 (3):195–205. https://www.ncbi.nlm.nih.gov/pubmed/19204717. 10.1038/nrg2526.
Commans, S., and A. Bock, 1999 Selenocysteine inserting tRNAs: an overview. *FEMS Microbiol Rev* 23 (3):335–351. https://www.ncbi.nlm.nih.gov/pubmed/10371037. 10.1111/j.1574-6976.1999.tb00403.x.
Crombie, T.A., R. McKeown, N.D. Moya, K.S. Evans, S.J. Widmayer *et al.*, 2024 CaeNDR, the Caenorhabditis Natural Diversity Resource. *Nucleic Acids Res* 52 (D1):D850–D858. https://www.ncbi.nlm.nih.gov/pubmed/37855690. 10.1093/nar/gkad887.
Crombie, T.A., S. Zdraljevic, D.E. Cook, R.E. Tanny, S.C. Brady *et al.*, 2019 Deep sampling of Hawaiian Caenorhabditis elegans reveals high genetic diversity and admixture with global populations. *Elife* 8. https://www.ncbi.nlm.nih.gov/pubmed/31793880. 10.7554/eLife.50465.
Danecek, P., J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan *et al.*, 2021 Twelve years of SAMtools and BCFtools. *Gigascience* 10 (2). https://www.ncbi.nlm.nih.gov/pubmed/33590861. 10.1093/gigascience/giab008.
Darrow, E.M., and B.P. Chadwick, 2014 A novel tRNA variable number tandem repeat at human chromosome 1q23.3 is implicated as a boundary element based on conservation of a CTCF motif in mouse. *Nucleic Acids Res* 42 (10):6421–6435. https://www.ncbi.nlm.nih.gov/pubmed/24753417. 10.1093/nar/gku280.
Davey-Young, J., F. Hasan, R. Tennakoon, P. Rozik, H. Moore *et al.*, 2024 Mistranslating the genetic code with leucine in yeast and mammalian cells. *RNA Biology* 21 (1):507–529. https://doi.org/10.1080/15476286.2024.2340297. 10.1080/15476286.2024.2340297.
Dittmar, K.A., J.M. Goodenbour, and T. Pan, 2006 Tissue-specific differences in human transfer RNA expression. *PLoS Genet* 2 (12):e221. https://www.ncbi.nlm.nih.gov/pubmed/17194224. 10.1371/journal.pgen.0020221.
Dolgin, E.S., B. Charlesworth, S.E. Baird, and A.D. Cutter, 2007 Inbreeding and outbreeding depression in Caenorhabditis nematodes. *Evolution* 61 (6):1339–1352. 10.1111/j.1558-5646.2007.00118.x.
Dowle, M., and A. Srinivasan, 2022 data.table: Extension of `data.frame`. https://Rdatatable.gitlab.io/data.table. https://Rdatatable.gitlab.io/data.table.
Duret, L., 2000 tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet* 16 (7):287–289. https://www.ncbi.nlm.nih.gov/pubmed/10858656. 10.1016/s0168-9525(00)02041-2.
FC, M., T.L. Davis, and ggplot2 authors, 2022 ggpattern: 'ggplot2' Pattern Geoms. https://CRAN.R-project.org/package=ggpattern. https://CRAN.R-project.org/package=ggpattern.
Félix, M.-A., R. Jovelin, C. Ferrari, S. Han, Y.R. Cho *et al.*, 2013 Species richness, distribution and genetic diversity of Caenorhabditis nematodes in a remote tropical rainforest. *BMC Evolutionary Biology* 13 (1):10. https://doi.org/10.1186/1471-2148-13-10. 10.1186/1471-2148-13-10.
Florentz, C., B. Sohm, P. Tryoen-Toth, J. Putz, and M. Sissler, 2003 Human mitochondrial tRNAs in health and disease. *Cell Mol Life Sci* 60 (7):1356–1375. https://www.ncbi.nlm.nih.gov/pubmed/12943225. 10.1007/s00018-003-2343-1.
Frézal, L., and M.-A. Félix, 2015 C. elegans outside the Petri dish. *Elife* 4:e05849. https://doi.org/10.7554/eLife.05849. 10.7554/eLife.05849.

Fusca, D.D., M.N. Dall'Acqua, S. Sánchez-Ramírez, and A.D. Cutter, 2025 Phylogenomic timetree-calibrated speciation clocks for Caenorhabditis nematodes reveal slow but disproportionate accumulation of post-zygotic reproductive isolation. *PLOS Genetics* 21 (9):e1011852. https://doi.org/10.1371/journal.pgen.1011852. 10.1371/journal.pgen.1011852.

Gaillard, H., and A. Aguilera, 2016 Transcription as a Threat to Genome Integrity. *Annu Rev Biochem* 85:291–317. https://www.ncbi.nlm.nih.gov/pubmed/27023844. 10.1146/annurev-biochem-060815-014908.

Galli, G., H. Hofstetter, and M.L. Birnstiel, 1981 Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* 294 (5842):626–631. https://www.ncbi.nlm.nih.gov/pubmed/7312050. 10.1038/294626a0.

Gao, L., A. Behrens, G. Rodschinka, S. Forcelloni, S. Wani *et al.*, 2024 Selective gene expression maintains human tRNA anticodon pools during differentiation. *Nat Cell Biol* 26 (1):100–112. https://www.ncbi.nlm.nih.gov/pubmed/38191669. 10.1038/s41556-023-01317-3.

Giege, R., M. Sissler, and C. Florentz, 1998 Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* 26 (22):5017–5035. https://www.ncbi.nlm.nih.gov/pubmed/9801296. 10.1093/nar/26.22.5017.

Goodenbour, J.M., and T. Pan, 2006 Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res* 34 (21):6137–6146. https://www.ncbi.nlm.nih.gov/pubmed/17088292. 10.1093/nar/gkl725.

Goto, Y., I. Nonaka, and S. Horai, 1990 A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature* 348 (6302):651–653. https://www.ncbi.nlm.nih.gov/pubmed/2102678. 10.1038/348651a0.

Gu, Z., L. Gu, R. Eils, M. Schlesner, and B. Brors, 2014 circlize Implements and enhances circular visualization in R. *Bioinformatics* 30 (19):2811–2812. https://www.ncbi.nlm.nih.gov/pubmed/24930139. 10.1093/bioinformatics/btu393.

Harris, C.R., K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen *et al.*, 2020 Array programming with NumPy. *Nature* 585 (7825):357–362. https://www.ncbi.nlm.nih.gov/pubmed/32939066. 10.1038/s41586-020-2649-2.

Hillier, L.W., R.D. Miller, S.E. Baird, A. Chinwalla, L.A. Fulton *et al.*, 2007 Comparison of C. elegans and C. briggsae genome sequences reveals extensive conservation of chromosome organization and synteny. *PLoS Biol* 5 (7):e167. https://www.ncbi.nlm.nih.gov/pubmed/17608563. 10.1371/journal.pbio.0050167.

Hipps, D., K. Shiba, B. Henderson, and P. Schimmel, 1995 Operational RNA code for amino acids: species-specific aminoacylation of minihelices switched by a single nucleotide. *Proc Natl Acad Sci U S A* 92 (12):5550–5552. https://www.ncbi.nlm.nih.gov/pubmed/7539919. 10.1073/pnas.92.12.5550.

Hughes, L.A., D.L. Rudler, S.J. Siira, T. McCubbin, S.A. Raven *et al.*, 2023 Copy number variation in tRNA isodecoder genes impairs mammalian development and balanced translation. *Nat Commun* 14 (1):2210. https://www.ncbi.nlm.nih.gov/pubmed/37072429. 10.1038/s41467-023-37843-9.

Iben, J.R., and R.J. Maraia, 2012 tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA* 18 (7):1358–1372. https://www.ncbi.nlm.nih.gov/pubmed/22586155. 10.1261/rna.032151.111.

Iben, J.R., and R.J. Maraia, 2014 tRNA gene copy number variation in humans. *Gene* 536 (2):376–384. https://www.ncbi.nlm.nih.gov/pubmed/24342656. 10.1016/j.gene.2013.11.049.

Ishimura, R., G. Nagy, I. Dotu, H. Zhou, X.L. Yang *et al.*, 2014 RNA function. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science* 345 (6195):455–459. https://www.ncbi.nlm.nih.gov/pubmed/25061210. 10.1126/science.1249749.

Jinks-Robertson, S., and A.S. Bhagwat, 2014 Transcription-associated mutagenesis. *Annu Rev Genet* 48:341–359. https://www.ncbi.nlm.nih.gov/pubmed/25251854. 10.1146/annurev-genet-120213-092015.

Kaur, T., and M.V. Rockman, 2014 Crossover heterogeneity in the absence of hotspots in Caenorhabditis elegans. *Genetics* 196 (1):137–148. https://www.ncbi.nlm.nih.gov/pubmed/24172135. 10.1534/genetics.113.158857.

Kim, C., J. Kim, S. Kim, D.E. Cook, K.S. Evans *et al.*, 2019 Long-read sequencing reveals intra-species tolerance of substantial structural variations and new subtelomere formation in C. elegans. *Genome Res* 29 (6):1023–1035. https://www.ncbi.nlm.nih.gov/pubmed/31123081. 10.1101/gr.246082.118.

Kobayashi, Y., M.Y. Momoi, K. Tominaga, T. Momoi, K. Nihei *et al.*, 1990 A point mutation in the mitochondrial tRNA(Leu)(UUR) gene in MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes). *Biochem Biophys Res Commun* 173 (3):816–822. https://www.ncbi.nlm.nih.gov/pubmed/2268345. 10.1016/s0006-291x(05)80860-5.

Kutter, C., G.D. Brown, A. Goncalves, M.D. Wilson, S. Watt *et al.*, 2011 Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nat Genet* 43 (10):948–955. https://www.ncbi.nlm.nih.gov/pubmed/21873999. 10.1038/ng.906.

Lant, J.T., M.D. Berg, I.U. Heinemann, C.J. Brandl, and P. O'Donoghue, 2019 Pathways to disease from natural variations in human cytoplasmic tRNAs. *J Biol Chem* 294 (14):5294–5308. https://www.ncbi.nlm.nih.gov/pubmed/30643023. 10.1074/jbc.REV118.002982.

Laricchia, K., S. Zdraljevic, D. Cook, and E. Andersen, 2017 Natural variation in the distribution and abundance of transposable elements across the Caenorhabditis elegans species. *Molecular Biology and Evolution* 34 (9):2187–2202.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan *et al.*, 2007 Clustal W and Clustal X version 2.0. *Bioinformatics* 23 (21):2947–2948. https://www.ncbi.nlm.nih.gov/pubmed/17846036. 10.1093/bioinformatics/btm404.

Lee, D., S. Zdraljevic, L. Stevens, Y. Wang, R.E. Tanny *et al.*, 2021 Balancing selection maintains hyper-divergent haplotypes in Caenorhabditis elegans. *Nat Ecol Evol* 5 (6):794–807. https://www.ncbi.nlm.nih.gov/pubmed/33820969. 10.1038/s41559-021-01435-x.

Leung, J., D.A. Sinclair, S. Hayashi, G.M. Tener, and T.A. Grigliatti, 1991 Informational redundancy of tRNA(4Ser) and tRNA(7Ser) genes in Drosophila melanogaster and evidence for intergenic recombination. *J Mol Biol* 219 (2):175–188. https://www.ncbi.nlm.nih.gov/pubmed/1903816. 10.1016/0022-2836(91)90560-s.

Li, C., W. Qian, C.J. Maclean, and J. Zhang, 2016 The fitness landscape of a tRNA gene. *Science* 352 (6287):837–840. https://www.ncbi.nlm.nih.gov/pubmed/27080104. 10.1126/science.aae0568.

Loewe, L., and A.D. Cutter, 2008 On the potential for extinction by Muller's ratchet in Caenorhabditis elegans. *BMC Evol Biol* 8:125. https://www.ncbi.nlm.nih.gov/pubmed/18447910. 10.1186/1471-2148-8-125.

Moya, N.D., S.M. Yan, R.C. McCoy, and E.C. Andersen, 2025 The long and short of hyperdivergent regions. *Trends Genet* 41 (4):303–314. https://www.ncbi.nlm.nih.gov/pubmed/39706705. 10.1016/j.tig.2024.11.005.

Murillo-Recio, M., M. Salvadores, A.G. Torres, L. Tsapanou, A.V. Picó *et al.*, 2025 Characterization of Somatic Mutations in Human tRNA Genes Reveals Tumor-Specific Mutational Loads, and the Generation of tRNA Variants that Alter the Genetic Code. *bioRxiv*:2025.2006.2009.658696. https://www.biorxiv.org/content/biorxiv/early/2025/06/09/2025.06.09.658696.full.pdf. 10.1101/2025.06.09.658696.

Murrell, P., 2014 gridBase: Integration of base and grid graphics. https://CRAN.R-project.org/package=gridBase. https://CRAN.R-project.org/package=gridBase.

Nawrocki, E.P., and S.R. Eddy, 2013 Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29 (22):2933–2935. https://www.ncbi.nlm.nih.gov/pubmed/24008419. 10.1093/bioinformatics/btt509.

Nei, M., and A.P. Rooney, 2005 Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121–152. https://www.ncbi.nlm.nih.gov/pubmed/16285855. 10.1146/annurev.genet.39.073003.112240.

Neuwirth, E., 2022 RColorBrewer: ColorBrewer Palettes. https://CRAN.R-project.org/package=RColorBrewer. https://CRAN.R-project.org/package=RColorBrewer.

Noble, L.M., J. Yuen, L. Stevens, N. Moya, R. Persaud *et al.*, 2021 Selfing is the safest sex for Caenorhabditis tropicalis. *Elife* 10. https://www.ncbi.nlm.nih.gov/pubmed/33427200. 10.7554/eLife.62587.

Ooms, J., 2025a magick: Advanced Graphics and Image-Processing in R. https://doi.org/10.32614/CRAN.package.magick. https://doi.org/10.32614/CRAN.package.magick.

Ooms, J., 2025b pdftools: Text Extraction, Rendering and Converting of PDF Documents. https://doi.org/10.32614/CRAN.package.pdftools. https://doi.org/10.32614/CRAN.package.pdftools.

Paradis, E., and K. Schliep, 2019 ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35 (3):526–528. https://www.ncbi.nlm.nih.gov/pubmed/30016406. 10.1093/bioinformatics/bty633.

Parisien, M., X. Wang, and T. Pan, 2013 Diversity of human tRNA genes from the 1000-genomes project. *RNA Biol* 10 (12):1853–1867. https://www.ncbi.nlm.nih.gov/pubmed/24448271. 10.4161/rna.27361.

Pedersen, T.L., 2022 ggforce: Accelerating 'ggplot2. https://CRAN.R-project.org/package=ggforce. https://CRAN.R-project.org/package=ggforce.

Perona, J.J., and I. Gruic-Sovulj, 2014 Synthetic and Editing Mechanisms of Aminoacyl-tRNA Synthetases, pp. 1–41 in *Aminoacyl-tRNA Synthetases in Biology and Medicine*, edited by S. Kim. Springer Netherlands, Dordrecht.

Pinzaru, A.M., and S.F. Tavazoie, 2023 Transfer RNAs as dynamic and critical regulators of cancer progression. *Nat Rev Cancer* 23 (11):746–761. https://www.ncbi.nlm.nih.gov/pubmed/37814109. 10.1038/s41568-023-00611-4.

R Core Team, 2025 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. https://www.R-project.org/.

Ren, K., and K. Russell, 2021 formattable: Create 'Formattable' Data Structures. https://CRAN.R-project.org/package=formattable. https://CRAN.R-project.org/package=formattable.

Rockman, M.V., and L. Kruglyak, 2009 Recombinational Landscape and Population Genomics of Caenorhabditis elegans. *PLOS Genetics* 5 (3):e1000419. https://doi.org/10.1371/journal.pgen.1000419. 10.1371/journal.pgen.1000419.

Rockman, M.V., S.S. Skrovanek, and L. Kruglyak, 2010 Selection at linked sites shapes heritable phenotypic variation in C. elegans. *Science* 330 (6002):372–376. https://www.ncbi.nlm.nih.gov/pubmed/20947766. 10.1126/science.1194208.

Rogers, H.H., C.M. Bergman, and S. Griffiths-Jones, 2010 The evolution of tRNA genes in Drosophila. *Genome Biol Evol* 2:467–477. https://www.ncbi.nlm.nih.gov/pubmed/20624748. 10.1093/gbe/evq034.

Rogers, H.H., and S. Griffiths-Jones, 2014 tRNA anticodon shifts in eukaryotic genomes. *RNA* 20 (3):269–281. https://www.ncbi.nlm.nih.gov/pubmed/24442610. 10.1261/rna.041681.113.

Ross, J.A., D.C. Koboldt, J.E. Staisch, H.M. Chamberlin, B.P. Gupta *et al.*, 2011 Caenorhabditis briggsae recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. *PLoS Genet* 7 (7):e1002174. https://www.ncbi.nlm.nih.gov/pubmed/21779179. 10.1371/journal.pgen.1002174.

Schimmel, P., 2018 The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol* 19 (1):45–58. https://www.ncbi.nlm.nih.gov/pubmed/28875994. 10.1038/nrm.2017.77.

Schliep, K.P., 2011 phangorn: phylogenetic analysis in R. *Bioinformatics* 27 (4):592–593. https://www.ncbi.nlm.nih.gov/pubmed/21169378. 10.1093/bioinformatics/btq706.
Schoenmakers, E., B. Carlson, M. Agostini, C. Moran, O. Rajanayagam *et al.*, 2016 Mutation in human selenocysteine transfer RNA selectively disrupts selenoprotein synthesis. *J Clin Invest* 126 (3):992–996. https://www.ncbi.nlm.nih.gov/pubmed/26854926. 10.1172/JCI84747.
Shih, D.J.H., 2021 argparser: Command-Line Argument Parser. https://CRAN.R-project.org/package=argparser. https://CRAN.R-project.org/package=argparser.
Sternberg, P.W., K. Van Auken, Q. Wang, A. Wright, K. Yook *et al.*, 2024 WormBase 2024: status and transitioning to Alliance infrastructure. *Genetics* 227 (1). https://www.ncbi.nlm.nih.gov/pubmed/38573366. 10.1093/genetics/iyae050.
Stevens, L., N.D. Moya, R.E. Tanny, S.B. Gibson, A. Tracey *et al.*, 2022 Chromosome-Level Reference Genomes for Two Strains of Caenorhabditis briggsae: An Improved Platform for Comparative Genomics. *Genome Biol Evol* 14 (4). https://www.ncbi.nlm.nih.gov/pubmed/35348662. 10.1093/gbe/evac042.
Sulston, J.E., and H.R. Horvitz, 1977 Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev Biol* 56 (1):110–156. https://www.ncbi.nlm.nih.gov/pubmed/838129. 10.1016/0012-1606(77)90158-0.
Surzycki, S.A., and W.R. Belknap, 2000 Repetitive-DNA elements are similarly distributed on Caenorhabditis elegans autosomes. *Proceedings of the National Academy of Sciences* 97 (1):245–249.
Suzuki, T., 2021 The expanding world of tRNA modifications and their disease relevance. *Nat Rev Mol Cell Biol* 22 (6):375–392. https://www.ncbi.nlm.nih.gov/pubmed/33658722. 10.1038/s41580-021-00342-0.
Suzuki, T., A. Nagao, and T. Suzuki, 2011 Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. *Annu Rev Genet* 45:299–329. https://www.ncbi.nlm.nih.gov/pubmed/21910628. 10.1146/annurev-genet-110410-132531.
Szymanski, M., A. Maurya, P. Kopec, and W.M. Karlowski, 2023 tRNA-Cys gene clusters exhibit high variability in Arabidopsis thaliana. *BMC Plant Biology* 23 (1):623. https://doi.org/10.1186/s12870-023-04632-x. 10.1186/s12870-023-04632-x.
Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123 (3):585–595. https://www.ncbi.nlm.nih.gov/pubmed/2513255. 10.1093/genetics/123.3.585.
Tennakoon, R., T.M.I. Bily, F. Hasan, K.S. Hoffman, and P. O'Donoghue, 2025 Natural human tRNAAla anticodon variants mistranslate the genetic code. *RNA* 31 (6):791–806. http://rnajournal.cshlp.org/content/31/6/791.abstract. 10.1261/rna.080450.125.
Thompson, O.A., L.B. Snoek, H. Nijveen, M.G. Sterken, R.J. Volkers *et al.*, 2015 Remarkably Divergent Regions Punctuate the Genome Assembly of the Caenorhabditis elegans Hawaiian Strain CB4856. *Genetics* 200 (3):975–989. https://www.ncbi.nlm.nih.gov/pubmed/25995208. 10.1534/genetics.115.175950.
Thornlow, B.P., J. Armstrong, A.D. Holmes, J.M. Howard, R.B. Corbett-Detig *et al.*, 2020 Predicting transfer RNA gene activity from sequence and genome context. *Genome Res* 30 (1):85–94. https://www.ncbi.nlm.nih.gov/pubmed/31857444. 10.1101/gr.256164.119.
Thornlow, B.P., J. Hough, J.M. Roger, H. Gong, T.M. Lowe *et al.*, 2018 Transfer RNA genes experience exceptionally elevated mutation rates. *Proc Natl Acad Sci U S A* 115 (36):8996–9001. https://www.ncbi.nlm.nih.gov/pubmed/30127029. 10.1073/pnas.1801240115.
Torres, A.G., 2019 Enjoy the Silence: Nearly Half of Human tRNA Genes Are Silent. *Bioinform Biol Insights* 13:1177932219868454. https://www.ncbi.nlm.nih.gov/pubmed/31447549. 10.1177/1177932219868454.
Torres, A.G., O. Reina, C. Stephan-Otto Attolini, and L. Ribas de Pouplana, 2019 Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. *Proc*

Natl Acad Sci U S A 116 (17):8451–8456. https://www.ncbi.nlm.nih.gov/pubmed/30962382. 10.1073/pnas.1821120116.
van den Brand, T., 2025 ggh4x: Hacks for 'ggplot2'. https://CRAN.R-project.org/package=ggh4x. https://CRAN.R-project.org/package=ggh4x.
Velandia-Huerto, C.A., S.J. Berkemer, A. Hoffmann, N. Retzlaff, L.C. Romero Marroquin *et al.*, 2016 Orthologs, turn-over, and remolding of tRNAs in primates and fruit flies. *BMC Genomics* 17 (1):617. https://www.ncbi.nlm.nih.gov/pubmed/27515907. 10.1186/s12864-016-2927-4.
Wang, P.P., and I. Ruvinsky, 2012 Family size and turnover rates among several classes of small non-protein-coding RNA genes in Caenorhabditis nematodes. *Genome Biol Evol* 4 (4):565–574. https://www.ncbi.nlm.nih.gov/pubmed/22467905. 10.1093/gbe/evs034.
Wickham, H., 2016 ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org. https://ggplot2.tidyverse.org.
Wickham, H., T. Pedersen, and D. Seidel, 2025 scales: Scale Functions for Visualization. https://CRAN.R-project.org/package=scales. https://CRAN.R-project.org/package=scales.
Wilke, C.O., 2020 cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. https://CRAN.R-project.org/package=cowplot. https://CRAN.R-project.org/package=cowplot.
Wolf, M., C. Koetschan, and T. Muller, 2014 ITS2, 18S, 16S or any other RNA - simply aligning sequences and their individual secondary structures simultaneously by an automatic approach. *Gene* 546 (2):145–149. https://www.ncbi.nlm.nih.gov/pubmed/24881812. 10.1016/j.gene.2014.05.065.
Yang, J., 2022 Tajima's D and genomic features. https://jyanglab.com/slides/2022-agro932/w6lab.html#2. https://jyanglab.com/slides/2022-agro932/w6lab.html#2.
Yarham, J.W., J.L. Elson, E.L. Blakely, R. McFarland, and R.W. Taylor, 2010 Mitochondrial tRNA mutations and disease. *Wiley Interdiscip Rev RNA* 1 (2):304–324. https://www.ncbi.nlm.nih.gov/pubmed/21935892. 10.1002/wrna.27.
Yona, A.H., Z. Bloom-Ackermann, I. Frumkin, V. Hanson-Smith, Y. Charpak-Amikam *et al.*, 2013 tRNA genes rapidly change in evolution to meet novel translational demands. *Elife* 2:e01339. https://www.ncbi.nlm.nih.gov/pubmed/24363105. 10.7554/eLife.01339.
Yu, G., 2022 Data Integration, Manipulation and Visualization of Phylogenetic Trees pp. 276. Chapman and Hall/CRC, New York. https://www.taylorfrancis.com/books/mono/10.1201/9781003279242/data-integration-manipulation-visualization-phylogenetic-trees-guangchuang-yu. 10.1201/9781003279242
Yu, G., 2025 scatterpie: Scatter Pie Plot. https://CRAN.R-project.org/package=scatterpie. https://CRAN.R-project.org/package=scatterpie.
Yu, G., D.K. Smith, H. Zhu, Y. Guan, and T.T.-Y. Lam, 2017 ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8 (1):28–36. https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12628. https://doi.org/10.1111/2041-210X.12628.