

Allelic Variation at tRNA Genes in Three Nematode Species Indicates Mutation Load Despite Strong Purifying Selection

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,*}

¹School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

²School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA

³Present address: AAAS, Washington, DC

⁴Present address: Department of Mathematics, California State University San Bernardino, San Bernardino, CA, USA

*Corresponding author: E-mail: paaby@gatech.edu.

Accepted: February 23, 2026

Abstract

Cytosolic transfer RNAs (tRNAs), which are encoded as hundreds of genes in nuclear genomes, experience exceptionally high mutation rates and have been hypothesized to confer substantial mutational load in natural populations. Although this phenomenon appears universal across multicellular eukaryotes, a comprehensive characterization of standing variation in tRNA repertoires is still lacking in any system. Here, 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 genes carry signatures of high rates of historical transcription-associated mutagenesis and of purifying selection, resulting in allelic variation that includes pervasive instances of within-gene mismatches between the amino acid recognized by the tRNA backbone and that indicated by the anticodon. Furthermore, patterns of tRNA genomic organization and variation differ markedly from those of protein-coding regions. Individual genomes harbor distinct complements of tRNA genes with predicted functional differences, an observation that coincides with recent evidence that variation in tRNA expression and regulation contributes to human disease. Our findings offer an entry point for identifying the microevolutionary processes that act on tRNA repertoires and, in turn, connecting those processes to the macroevolutionary patterns that have more frequently been the focus of study.

Key words: mutation, mutation load, *C. elegans*, *Caenorhabditis*, isotype switching, transcription-associated mutagenesis.

Significance

Transfer RNAs (tRNAs) are ancient and essential molecules that deliver amino acids to the growing polypeptide chain during protein synthesis. Despite their importance, little is known about how the genes encoding tRNAs vary among individuals, even though they accumulate large numbers of mutations due to their high rates of expression. Here, we report that three nematode species, including the model organism *C. elegans*, show within-species variation in their tRNA gene repertoires, which may cause differences in protein synthesis across individuals. Our findings may represent patterns common across multicellular species, and they set the stage for future investigation into how this diversity influences cellular function and organismal fitness.

© The Author(s) 2026. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

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 deep conservation of the stereotypical “cloverleaf” secondary structure (Fig. 1a) and L-shaped tertiary structure integral to their molecular function. tRNAs are specific to their cognate amino acid, typically occurring as one of 22 isotypes (corresponding to the 20 canonical amino acids plus two modified amino acids). The isotype of a tRNA is determined by its anticodon, which recognizes codons in the mRNA, as well as other sequence identity elements necessary for the binding specificity of the appropriate aminoacyl-tRNA synthetase; these may reside in the acceptor stem, for example (Fig. 1a; Giege and Eriani 2023).

However, despite ancient conservation of this molecular structure–function relationship, tRNA sequences in nuclear genomes undergo extensive evolutionary turnover. These cytosolic tRNAs typically occur as hundreds of genes in eukaryotic genomes, often including dozens per isotype (Goodenbour and Pan 2006; Bermudez-Santana et al. 2010). Across eukaryotic lineages, cytosolic 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) (Green et al. 2003; Jinks-Robertson and Bhagwat 2014; Thornlow et al. 2018), which may include as-yet uncharacterized mechanisms associated with Pol III transcription in particular (Seplyarskiy et al. 2023). Like other aspects of tRNAs, this phenomenon appears universal: polymorphism at tRNA loci in humans, *Mus musculus*, *Drosophila melanogaster*, and *Arabidopsis thaliana* shows strong signatures of TAM (Thornlow et al. 2018). During tRNA transcription, flanking regions just upstream and downstream of the gene are likewise subject to TAM. These flanking

regions exhibit dramatically elevated rates of substitution relative to the gene body, providing a record of historical exposure to TAM and indicating that functional tRNA sequences are subject to strong purifying selection (Thornlow et al. 2018). Modeling the TAM signature via a sequence substitution model, we find that standing variation in *C. elegans* tRNAs is well explained by historical exposure to TAM (Banos et al. 2025).

Thornlow and colleagues (2018) have proposed that the elevated mutation rate at nuclear-encoded tRNAs may confer a significant mutation load on eukaryotic genomes. Exploration of this hypothesis ultimately requires evaluating the fitness consequences of standing genetic variation in tRNAs within extant populations. Mutations to mitochondrial tRNAs have long been known to cause an array of diseases (Goto et al. 1990; Kobayashi 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. At nuclear-encoded tRNAs, tRNA gene redundancies make it difficult to identify, and empirically test, variants with functional consequences, though mutations in human cytosolic tRNAs predicted to cause amino acid mis-loading can induce biochemical consequences (Davey-Young et al. 2024; Tennakoon et al. 2025). However, recent studies indicate growing interest in allelic variation of cytosolic tRNAs, especially in the context of disease (Schoenmakers et al. 2016; Thornlow et al. 2018; Lant et al. 2019; Pinzaru and Tavazoie 2023; Murillo-Recio et al. 2025). In humans, studies with small sample sizes reveal substantial variation at tRNA genes, even as they likely under-sample the breadth of variation (Darrow and Chadwick 2014; Iben and Maraia 2014; Berg et al. 2019). Larger-scale, lower-coverage sequencing also indicates extensive variation (Parisien et al. 2013; Lant et al. 2019), though these approaches do not capture the rare variants that likely contribute disproportionately to functional differences. The consensus view holds that the full extent of cytosolic tRNA variation remains underestimated (Lant et al. 2019).

In parallel, the rapidly emerging field of tRNA regulation has demonstrated a link between the dysregulation of cytosolic tRNAs and disease. These findings, facilitated by new approaches to analyzing tRNA expression, show that tRNA regulatory control is complex and dynamic across age, cell type, and environmental cues (Suzuki 2021; Pinzaru and Tavazoie 2023). These studies directly connect cytosolic tRNA activity to organismal fitness, though so far this research area has largely overlooked the consequence of mutational variation on posttranscriptional tRNA function.

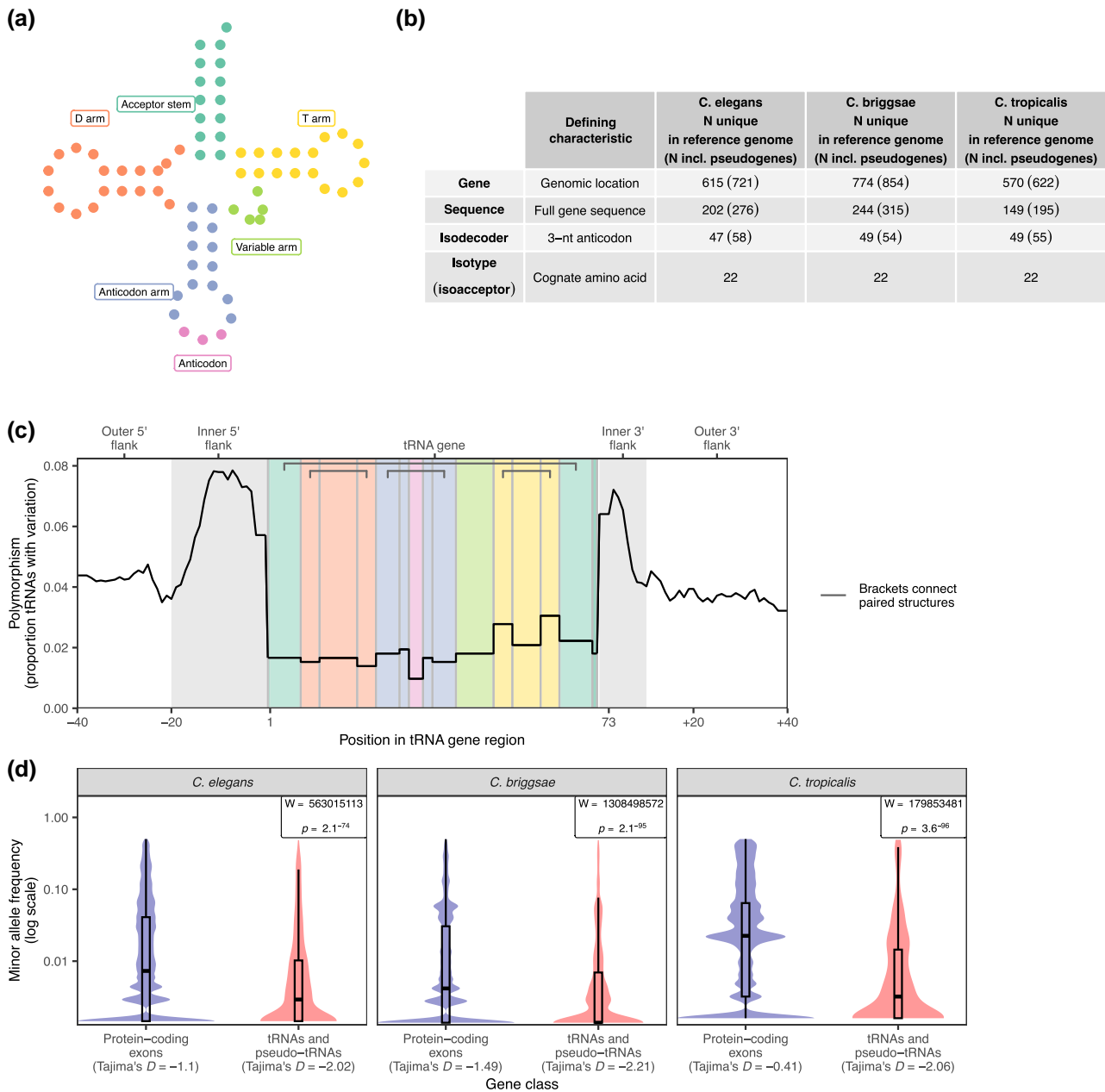


Fig. 1. Structural, functional, sequence similarity, and transcription-associated mutagenesis (TAM) of tRNA genes in *Caenorhabditis* nematodes. a) “Cloverleaf” secondary structure of a canonical 72 bp tRNA, with each base plotted as a circle and colored and labeled by the component of the tRNA structure 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: individual genes (which have distinct genomic locations), distinct full-tRNA sequences (which may occur at multiple genomic locations), distinct anticodons (isodecoders), and distinct amino acids (isotypes/isoacceptors). Numbers outside of parentheses represent putatively functional tRNA genes, while numbers in parenthesis include these as well as predicted nonfunctional tRNA 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 MAF < 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) MAF spectrum of variants (SNVs and INDELS) in protein-coding exons and putatively functional tRNA genes; 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.

To begin to evaluate the fitness consequences of mutational variation in cytosolic tRNAs, we first require comprehensive characterization of tRNA allelic variation at the population level. The universality of the genetic code, deep conservation of translational machinery, and ubiquity of TAM-induced substitutions at tRNA loci (Thornlow et al. 2018) indicate that insights from any system should be informative across eukaryotes. For example, despite their differences in organismal complexity, the genomes of flies and worms to humans and chimps encode comparable number of tRNAs (Goodenbour and Pan 2006), which vary by isotype copy number and reflect adaptation to codon usage (Duret 2000; Qian et al. 2012). Here, we evaluate genotype data for hundreds of wild strains in three species of *Caenorhabditis* nematode worms: the model organism *C. elegans* and related species *C. briggsae* and *C. tropicalis* (Crombie et al. 2024). These species diverged from each other ~10 to 25 million years ago and are all androdioecious, i.e. hermaphroditic and largely self-fertilizing in nature (Stevens et al. 2022; Fusca et al. 2025). This mating system may permit 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), enabling the detection of rare variants and alleles.

In this study, we resolved within-species allelic variation in nuclear-encoded tRNAs from >620 wild strains in each of the three *Caenorhabditis* species. Then, we asked (i) what evolutionary forces shape these tRNA repertoires, (ii) to what extent the observed allelic variation may affect tRNA function, including via within-gene isotype mismatches between the tRNA backbone (scaffold) and the anticodon, and (iii) whether the genomic organization of tRNA genes departs from that of protein-coding regions and that predicted by genome architecture.

Results

tRNA Repertoires in *Caenorhabditis*

To characterize the pool of tRNA genes encoded in the nuclear genomes of *C. elegans*, *C. briggsae*, and *C. tropicalis*, we first computationally predicted tRNA-like sequences in the three reference genomes. As tRNAs exhibit highly conserved “cloverleaf” secondary structure and corresponding genomic sequence motifs (Fig. 1a), putatively functional tRNAs and nonfunctional tRNA-like sequences (pseudogenes) can be identified with high confidence from sequence data (Chan et al. 2021). Each tRNA gene belongs to multiple nested classes (Fig. 1b): the gene itself defined by its unique

location, a set of one or more genes with identical sequence, an isodecoder class of genes that share the same anticodon and therefore cognate amino acid, and an isotype (also called isoacceptor) class that shares the same cognate amino acid but not necessarily the same anticodon. Each species encoded tRNAs for all 20 canonical amino acids as well as initiation-specific methionine (iMet) and selenocysteine (SeC); the total number of putatively functional tRNA genes within each reference genome was 615 (*C. elegans*), 774 (*C. briggsae*), and 570 (*C. tropicalis*) (Fig. 1b). Within species, these genes often shared sequence identity, such that the number of unique tRNA sequences ranged from 149 to 244 per reference genome, in turn comprised of 48 to 49 unique anticodon sequences (Fig. 1b, Tables S1 to S2). These patterns were consistent with known tRNA gene complements (Chan and Lowe 2009, 2016).

We first investigated isotype and isodecoder frequency at tRNA genes. While copy number varied substantially, with some isotypes differing by a third or more across species (Table S1), overall isotype frequencies were broadly similar (Pearson's $R^2 \geq 0.92$, $P < 4.3 \times 10^{-13}$ for gene number across isotypes). For example, glycine and leucine rank first and second in isotype prevalence in *C. elegans* and *C. briggsae*; in *C. tropicalis*, they rank third and first, with proline the second most prevalent isotype. These differences represent relative isotype representation shifts of up to 30% across species, though the individual isotype frequencies are modest within the full repertoire (e.g. a shift in proline frequency from 0.068 in *C. elegans* to 0.089 in *C. tropicalis*). In all three species, selenocysteine (SeC) was the rarest tRNA encoded amino acid (Table S1), consistent with the rarity and unique role of SeC across the tree of life (Commans and Bock 1999; Chan and Lowe 2016).

At the codon level, most codon sequences (corresponding to the anticodon of the transcribed tRNA molecule) appeared in multiple tRNA genes within each reference genome. 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). Each reference genome contained tRNA genes with a single instance of a given codon (three in *C. elegans*, five in *C. briggsae*, two in *C. tropicalis*), though these were not necessarily the same codons, and several (out of 61 possible) were entirely absent. Missing isodecoders are a persistent feature of tRNA repertoires in all living systems (Chan and Lowe 2016), even as the codons themselves appear in protein-coding genes throughout the genome; wobble pairing at the third position permits codon reading by non-cognate tRNAs (Crick 1966; Roth 2012). Missing isodecoders tended to be consistent across the

three species: of the 18 codon sequences missing in any species, 13 were missing in all three (Table S2).

Caenorhabditis tRNA Genes Exhibit Signatures of Transcription-Associated Mutagenesis

tRNAs in humans, mice, *Arabidopsis*, and *Drosophila* show evidence of exceptionally high rates of mutation, incurred because of their high rate of transcription and evidenced by patterns of intraspecific polymorphism (Thornlow et al. 2018). To assess the role of TAM in our system, we examined population variation at tRNA loci and their flanking genomic regions in our three species. During transcription, a gene and its flanking regions are unwound, and the unbound, non-template strand is vulnerable to mutation. For tRNAs, the elevated rate of transcription induces an exceptionally high mutation rate; a substantial fraction of the mutations are purged from the gene body by purifying selection, but not from the flanking regions, which provide a historical record of mutagenic activity (Jinks-Robertson and Bhagwat 2014; Thornlow et al. 2018). Consistent with the reports in other systems (Thornlow et al. 2018) and our earlier findings in *C. elegans* (Banos et al. 2025), tRNA genes and their flanking regions exhibited signatures of TAM in the genetic variation of wild strains in all three species (Fig. 1c, Figs. S1 and S2).

Specifically, tRNA gene flanking regions had exceptionally high mutational variation, with up to 7.6% to 12.9% of tRNA genes exhibiting mutations at a given base across species (Fig. 1c, Fig. S1). The substitutions most often induced by TAM, G>A and C>T (Jinks-Robertson and Bhagwat 2014; Gaillard and Aguilera 2016), tended to be the most frequent mutations in both upstream tRNA gene flanking regions and within tRNA gene bodies (Fig. S2). Consistent with strong purifying selection, tRNA gene bodies exhibited many fewer mutations than flanking regions (Fig. S1). Nevertheless, we observed mutations across the whole length of the tRNA in each of the three species (Fig. S1). As we observed in a *C. elegans*-specific analysis (Banos et al. 2025), polymorphism was elevated at paired stem sites relative to loops. This is likely driven by increased susceptibility to TAM arising from the higher GC content at paired sites (Banos et al. 2025), though it is also consistent with stronger purifying selection on loops, which indicate greater fitness costs when mutated (Li et al. 2016).

Under the assumption that the rate of deleterious mutation and subsequent purifying selection is exceptionally high for tRNAs, we hypothesized that tRNA genes might be enriched for low-frequency variants relative to other functionally important elements.

Indeed, compared to protein-coding genes, variants in putatively functional tRNA genes tended to have lower minor allele frequencies (Fig. 1d), and these variants were much more likely to be present in the population as singletons (49% to 53% singleton variants in tRNAs vs. 24% to 37% in protein-coding regions across species, Table S3). This excess of rare alleles was likewise reflected in lower estimates of Tajima's *D* (Fig. 1d; Table S4), which assesses site frequency spectra by comparing pairwise differences to the number of segregating sites (Tajima 1989). The difference between protein-coding genes and tRNAs remained strongly significant for each species comparison when all tRNAs, including those predicted to be nonfunctional, were included, though nonfunctional sequences had higher minor allele frequencies and smaller singleton proportions (Table S3). This shift likely reflects stronger purifying selection on active tRNA loci.

Allelic Variation at *Caenorhabditis* tRNA Genes With Putatively Functional Consequences

Next, we sought to characterize intraspecific allelic variation at tRNA genes within the three *Caenorhabditis* species. To do this, we reconstructed full-length alleles for each gene in each wild strain (>620 per species), applying variant genotype data from CaENDR (Crombie et al. 2024) to the reference genomes' tRNA gene sets. Some tRNA genes were invariant, exhibiting only one allele; variable genes harbored up to 17 alleles across the populations (Fig. 2, Fig. S3). Among putatively functional tRNA genes, 38.8% to 83.3% showed allelic variation across the three species (Fig. S3). Across all nuclear-encoded tRNAs, including putative pseudogenes predicted to be nonfunctional, individual tRNA isotypes comprised 1 to 139 alleles in *C. elegans* (of 1695 total unique alleles at all tRNA genes in the entire population), 4 to 328 in *C. briggsae* (of 3574 total), and 1 to 99 in *C. tropicalis* (of 1048 total).

The extensive allelic diversity over so many tRNA genes resulted in a unique complement of putatively functional tRNA alleles for each strain of *C. elegans* and *C. briggsae*. In *C. tropicalis*, which has lower average genetic diversity (Noble et al. 2021), 87% of strains (545 of 622) exhibited unique tRNA allelic complements; the remaining strains shared tRNA allele identity in sets of two or three strains, with the exception of one allelic combination, which was shared across nine strains.

To investigate potential functional relevance of this allelic variation, we developed allele classifications by combining information about: the predicted amino acid for that tRNA, inferred from the tRNA backbone sequence excluding the anticodon; the amino acid matching the anticodon; and the estimated tRNA folding

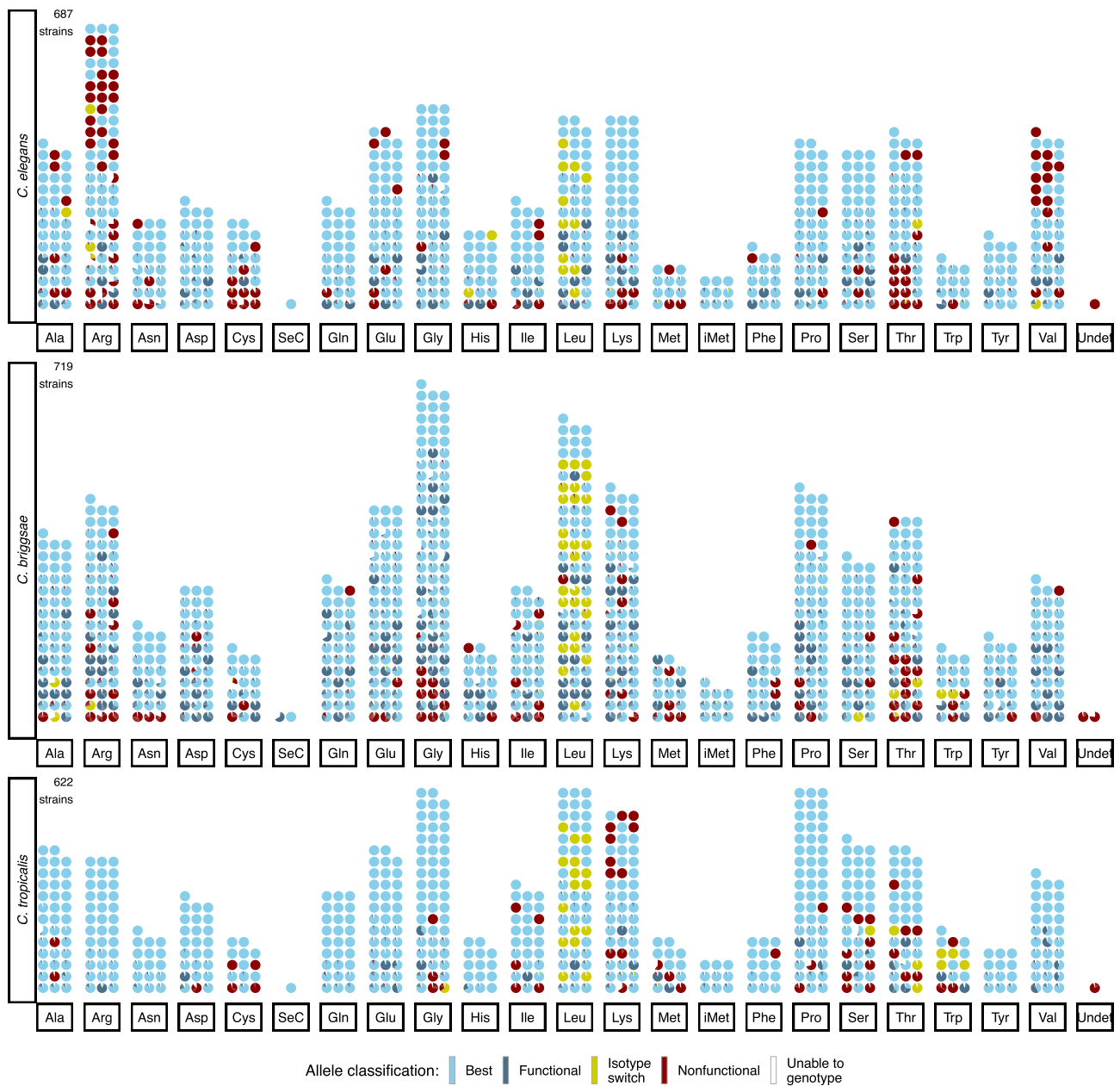


Fig. 2. Allelic diversity at all tRNA genes in *Caenorhabditis* reference genomes. Each tRNA gene in the reference genome of each species is represented by a circle, grouped by the amino acid (isotype) predicted to match its backbone; slices of the circle correspond to different alleles of that gene (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 a searchable web app at <https://wildworm.biosci.gatech.edu/trnarefalleles/>.

score (all from tRNAscan-SE). Alleles for which the anticodon and backbone predicted 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 were classified as “Nonfunctional” by low folding scores or non-identifiable anticodons, and “Isotype switch” by a mismatch between the amino acid predicted from the backbone sequence and the amino acid matching the anticodon; isotype switching has long been observed

across tRNA evolution at interspecies timescales (Rogers and Griffiths-Jones 2014).

Within each species, a substantial proportion of the nuclear tRNA genes exhibited putatively functional allelic diversity, with alleles belonging to all categories (Fig. 2). Of the genes classified as “Nonfunctional” in the reference genomes (Fig. 1b), >91% in each species exhibited only “Nonfunctional” alleles (or missing data) across the strains. However, genes in all three species harbor both functional and nonfunctional alleles (46 in *C. elegans*, 99 in *C. briggsae*, and 10 in *C. tropicalis*), suggesting that some strains in the population express viable tRNAs for that specific gene while others do not (Fig. 2). tRNA isotypes varied in the degree of predicted functional allelic diversity, and not strictly by 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 (Fig. 2).

We observed several instances in which the reference genome allele was a poor representative of the overall population. For example, three genes in *C. briggsae*, one in *C. elegans*, and one in *C. tropicalis* harbored alleles observed only in the reference strain, each of which was suboptimal relative to other alleles (with an RNA folding score up to 20% lower than that of the best allele). This observation emphasizes the relevance of the population as a whole when defining tRNA repertoires. Relatedly, the occurrence of putatively functional allele(s) throughout the population at a given locus increases confidence that the gene is maintained by selection. We examined cases of rare isodecoders in which the gene set derived from the reference genome included a single instance of a codon in one species that was absent from the other two species. Four such instances occurred, as well as a fifth in which the codon was present in two of the three species (Table S2). While we hypothesized that these rare codons might reflect reference strain-specific mutations absent from the broader population, in fact, all strains harbored predicted functional alleles—some of which were fixed in the population—of the same isodecoder class at these loci. We also examined the two SeC tRNAs in *C. briggsae*; most organisms encode a single nuclear SeC (Chan and Lowe 2016). One SeC gene was invariant across the population, as in *C. elegans* and *C. tropicalis*, while the other gene had three observed alleles and exhibited missing genotypes in one third of the strains. The rare isodecoder tRNAs and alleles of the second SeC gene exhibited lower RNA folding scores relative to similar genes, raising questions about their biological function, even as their apparent persistence in the population motivates investigation into their origin and potential maintenance.

Alleles with apparent isotype switches might be presumed deleterious, given their potential to misload amino acids (Giege et al. 1998). Correspondingly, of genes with isotype switch alleles that were not fixed in the population, 70% to 100% exhibited switch alleles at very low frequency (<0.5%) in the three *Caenorhabditis* species (Fig. 2, small yellow pie pieces). Alternatively, some isotype switches are tolerated and can persist across taxa (Yona et al. 2013; Rogers and Griffiths-Jones 2014). We also observe many such isotype switch alleles at high frequency and fixed within species, perhaps reflecting ancient switches (Fig. 2, fully or nearly fully yellow pies). We next investigated these observations further.

Evidence for Both Contemporaneous and Ancient Isotype Switching in *Caenorhabditis*

We examined isotype switching in the three species, both to assess polymorphism of isotype switching within extant populations and identify potential ancestral changes that may have arisen in a common ancestor of one or more *Caenorhabditis* species (~25 mya). An “isotype switch” allele may arise by mutation to the anticodon or by alteration of the backbone sequence, either of which may induce a mismatch in the identity of the associated amino acid via even a single nucleotide change (Hipps et al. 1995). When isotype switching was variable in the population, we could confidently infer the derived variant(s) in the majority of cases, such that we continued with this analysis (Fig. S4, Methods).

We characterized isotype switching in terms of the frequency of isotype switching from each amino acid, as well as the frequency of the isotype switch allele(s) themselves (Fig. 3a). Genes with isotype switch alleles fell into three major categories: (i) “all switch”: all alleles were isotype switches (either one or multiple alleles with the same switch); (ii) “no switch”: no alleles exhibited predicted isotype mismatches; and (iii) “some switch, some not”: genes with a mix of switching and non-switching alleles (Fig. 3a). Isotype switches that were fixed within a population 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 (Fig. 3a; Table S5). The incidence of isotype switching was not well predicted by 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 the near-highest gene count but only exhibited low-frequency variable switching in *C. briggsae* (Fig. 3a). Per species, the degree of fixed isotype switches, presumably arising deeper in time from a common

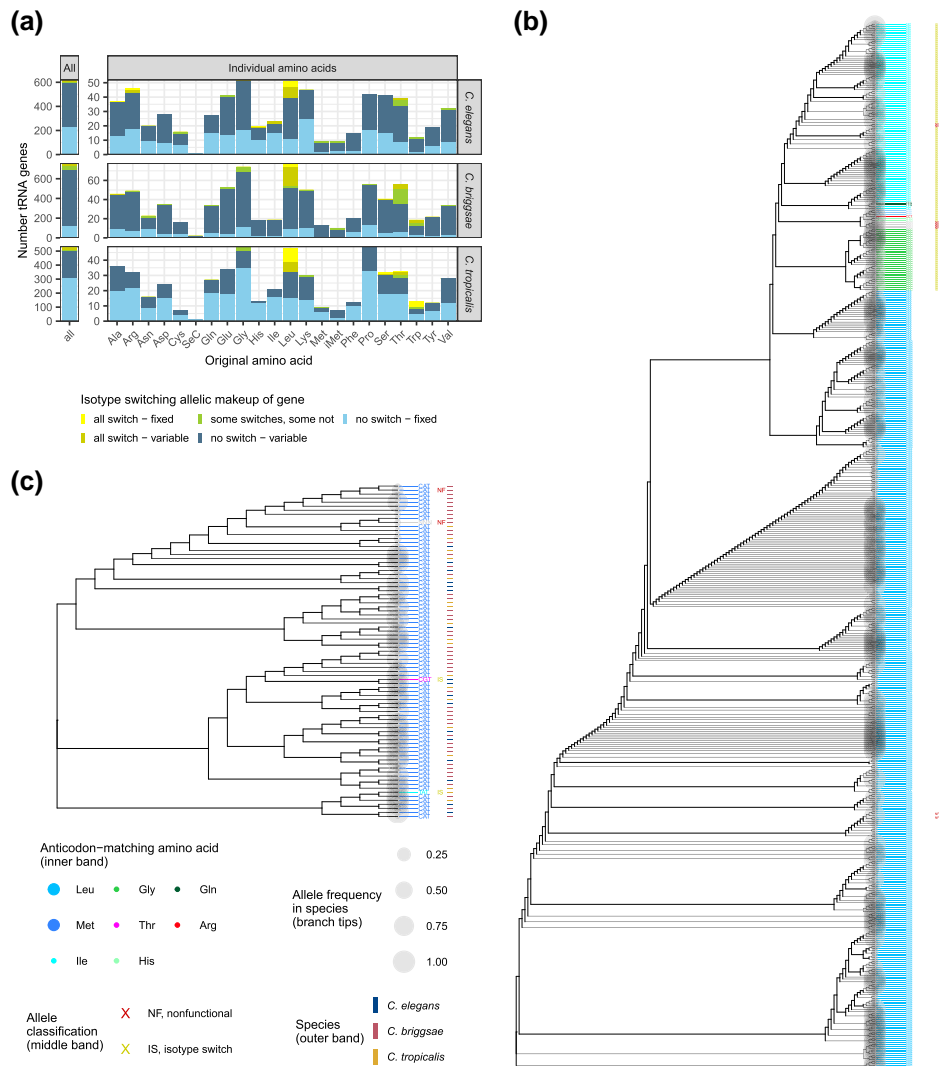


Fig. 3. Isotype switching from ancient and contemporaneous tRNA genetic variation. a) The number of tRNA genes predicted to encode each amino acid, colored by whether the gene’s anticodon matches the prediction and whether there is allelic variation at that gene in the species. Color definitions: all switch—fixed: a single isotype-switching (anticodon-backbone mismatch) allele is present in all strains; all switch—variable: all alleles harbor isotype switches, but there are multiple alleles in the population; some switches, some not: some alleles observed in the population harbor isotype switches while others do not (variable isotype switching); no switch—variable: no alleles in the population harbor isotype switches, but there are multiple alleles in the population; no switch—fixed: one allele without an isotype switch is fixed in the population. b, 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 nonfunctional 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 matches threonine (magenta) in a minority of strains in *C. elegans* and isoleucine (cyan) in a minority of strains in *C. briggsae*.

ancestor, did not correlate with species-wide measures of genetic or tRNA-specific variation: *C. elegans* (intermediate variation) had seven amino acids with fixed switches, *C. briggsae* (highest variation) had five, and *C. tropicalis* (lowest variation) had four (Fig. 3a), corresponding to 19, 31, and 25 isotype-switched tRNA genes, respectively.

To further evaluate fixed isotype switches, 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, genes with leucine backbones 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 (Fig. 3b). Similarly, another set of alleles with leucine backbones showed switches to glycine in all three species and clustered together across species (Fig. 3b). These examples support the conclusion that some switches arose from ancient, shared events and may be neutral or beneficial. Notably, one gene with a leucine backbone (*C. elegans* l.trna17) exhibited isotype switch alleles for every strain, but 80.2% of the strains carried alleles with a glutamine-recognizing anticodon (TTG in the DNA) while the remaining 19.8% carried alleles with anticodons for isoleucine (TAT).

In genes polymorphic for switching (“some switch, some not”), the switches tended to be species-specific and low frequency, possibly reflecting relatively recent, deleterious mutations that may be under purifying selection (Fig. 3a, Table S6). The tRNAs for methionine exemplified this trend (Fig. 3c). Methionine switched to threonine in a singleton (observed in one strain) allele in *C. elegans* and to isoleucine in a singleton allele in *C. tropicalis* (Fig. 3c). These rare alleles did not cluster closely on the allele tree, suggesting they arose as independent, relatively recent events (Fig. 3c). 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 (Fig. 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 (Fig. 1c).

We observed only one gene in which isotype switching was neither rare nor fixed (*C. elegans* tRNA gene V.trna22) (Table S6). This backbone sequence of this gene predicted valine and, for 546 strains (79.7% of the population), exhibited an anticodon corresponding to methionine (CAT), including the reference strain. The remaining 118 strains (12.3%) carried alleles without an isotype switch and encoded a valine anticodon (CAC) that matches the backbone. This example highlights the potential for studying isotype switching in *Caenorhabditis* species, from ancient events to contemporaneous variation.

Distinctive Genomic Organization of tRNA Genes

Genomic context can provide important clues about evolutionary processes. Therefore, we examined how tRNA genes were organized relative to each other and within the architecture of *Caenorhabditis* genomes. For these analyses, we used the positions of tRNA genes derived from the reference genomes.

We first asked whether the distribution of tRNA genes across *Caenorhabditis* genomes exhibited the

same patterns as protein-coding genes. The holocentric chromosomes of these species exhibit distinct recombination domains, with higher recombination rates in chromosome arms and lower rates 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, overall gene density is generally higher and genetic variation lower in centers relative to 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 (Fig. 4a). However, tRNA genes did not follow this rule: they were distributed more evenly across chromosomes, sometimes even enriched in chromosome arms (Fig. 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 in each species (Fig. 4a).

Likewise, we examined the incidence of tRNA genes in hyperdivergent regions of the *C. elegans* genome, in which the sequence is highly differentiated in at least one strain from other haplotypes in the population (Lee et al. 2021) (Table S7). While ~24% of *C. elegans* protein-coding genes occur in such intervals, only ~15% of putatively functional tRNA genes do ($\chi^2 = 33.4$, $d.f. = 1$, $P = 2.3 \times 10^{-8}$). However, nonfunctional tRNAs, which we presume are pseudogenized, were significantly more likely than protein-coding genes to reside in hyperdivergent regions, at 39% ($\chi^2 = 11.4$, $d.f. = 1$, $P = 0.002$). This observation prompts questions about the relevant timescale for tRNA “turnover” as well as the role of selection and hypermutability in evolutionary trajectories of both tRNAs and hyperdivergent regions (Lee et al. 2021; Moya et al. 2025a; Moya et al. 2025b).

In all three species, tRNA genes occurred on all chromosomes in all center and arm recombination domains (Fig. 4, Fig. S5). All isotypes comprising more than one tRNA gene had genes on at least half of all chromosomes (three of six) (Fig. 4, Fig. S5). Given this genome-wide distribution, we next asked whether similar tRNA genes were more likely to reside near each other, or if gene locations were random with regard to identity at the isotype, isodecoder, or full sequence level. tRNAs showed evidence of a dispersed, nonstereotyped genomic distribution, as numerous genes with shared identity, including identical sequences, occurred on different chromosomes. For example, in each species, nine genes for initiation methionine (iMet) were found on three (*C. tropicalis*) or four (*C. elegans*, *C. briggsae*) chromosomes. In *C. elegans* and *C. tropicalis*, all nine iMet tRNA genes had identical sequences in the reference genome; in *C. briggsae*, two distinct sequences corresponded to iMet. Other isotypes, such as valine (Fig. 4b, Fig. S5),

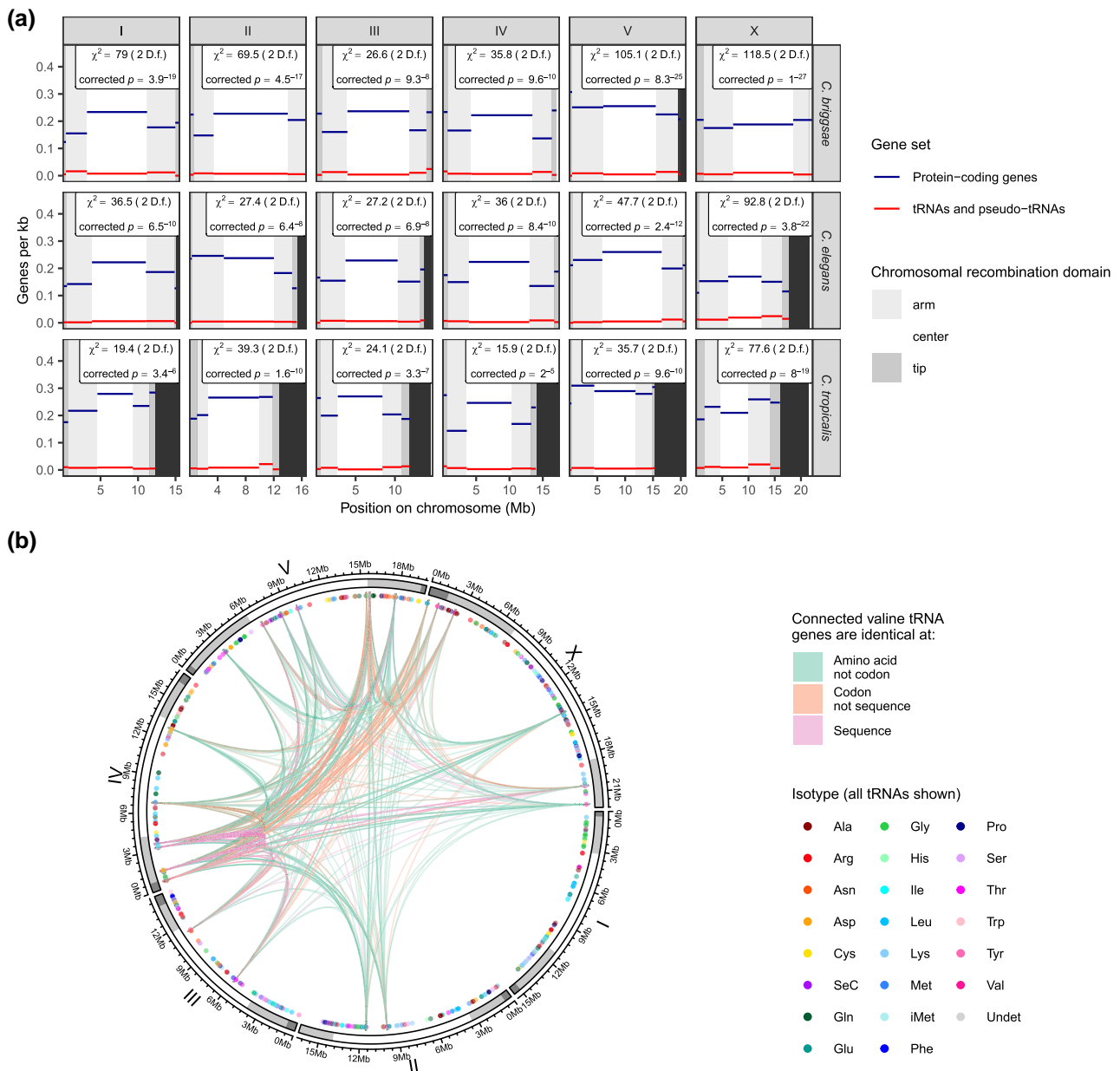


Fig. 4. Distribution of tRNA genes within *Caenorhabditis* genomes. a) Per-kb density of protein-coding and tRNA genes in each chromosome's high recombination (arms) and low recombination (center) domains. Species' chromosomes have different lengths; where one species has a longer chromosome than another, 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 \times 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 the chromosomal recombination domain (see a).

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 reside on the same chromosome

the more similar they were: genes within an isotype were more likely to share a 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 (Fig. S6). Furthermore, within the same chromosome, tRNA gene pairs tended to be closer together the more similar they were. This pattern followed the same hierarchy as above and also showed that sequence similarity correlated with genomic proximity (Fig. S7).

Discussion

The universality of tRNAs, including their essentiality, organization in eukaryotic nuclear genomes, and vulnerability to TAM, motivates a comprehensive, high-resolution characterization of their allelic variation—in any representative system. *Caenorhabditis* nematodes offer an ideal opportunity, including hundreds of genetically diverse strains and extensive genetics tools developed for the model organism *C. elegans*, which will maximize tractability in follow-on experiments (Barriere and Felix 2005a, 2005b; Félix et al. 2013; Crombie et al. 2019; Andersen and Rockman 2022; Crombie et al. 2024).

Perspective

Our nascent understanding of natural genetic variation in tRNA repertoires, which this work aims to advance, points to unresolved questions that are both broad and biologically significant. For one, does standing variation at tRNA loci confer functional differences among individuals and costs to fitness (Thornlow et al. 2018)? Here, we show that the tRNA pools of three *Caenorhabditis* species harbor extensive intra-specific variation in tRNA genes, including substitutions at every site along the generic molecule, frequency spectra consistent with deleterious effects, and derived mismatches between the amino acid decoded by the anticodon and that specified by other identity elements. In a parallel study, in which we introduced a novel sequence substitution model for TAM, we demonstrated that standing mutations in *C. elegans* cytosolic tRNAs are consistent with the TAM signature and destabilize molecular structure by disproportionately affecting paired sites and reducing thermodynamic stability (Banos et al. 2025). These findings support the hypothesis that tRNA loci contribute meaningful mutational load, though experimental assays will be required to test how these mutations affect molecular and organismal fitness.

Mutations at tRNAs that occur as single-copy genes (e.g. tRNAs for selenocysteine and mt-tRNA) demonstrate strong deleterious effects (Florentz et al. 2003; Yarham et al. 2010; Suzuki et al. 2011; Schoenmakers et al. 2016; Lant et al. 2019), but the high copy number and multidimensional variation of nuclear encoded tRNAs add

significant biological and analytical complexity. Nevertheless, recent work in mammals (Torres 2019; Hughes et al. 2023; Gao et al. 2024) and yeast (Bloom-Ackermann et al. 2014) suggests that specific genes within isoacceptor and isodecoder classes exert disproportionate functional impact, despite apparent functional redundancy (Leung et al. 1991; Kutter et al. 2011). An essential follow-up to the current study will be to intersect polymorphism data at the population level with inferences about tRNA activity. tRNA activity may be estimated from expression or chromatin accessibility data; computational models that integrate these data sources with estimates of historical expression levels (from TAM-induced mutational variation in flanking regions) and canonical tRNA functionality (e.g. (Thornlow et al. 2020)) offer a promising path for inferring functional consequences of observed mutations. tRNAs are also among the most decorated molecules for posttranscriptional modification, which plays a central role in their regulation and function; so incorporating this dimension will be important (Schimmel 2018; Torres 2019; Suzuki 2021; Davey-Young et al. 2024; Tennakoon et al. 2025). tRNA gene variants may themselves alter expression, as many tRNAs have internal promoters (Galli et al. 1981; Giege et al. 1998); they are also likely to affect posttranscriptional regulation and activity, including by posttranscriptional modification (Schimmel 2018). Direct quantification and epigenetic profiling, using methods that distinguish identical genes via flanking sequence awareness, in individuals with diverse allelic repertoires and across multiple tissues and life stages, will be informative.

Another question arising from intraspecific variation at tRNA genes relates to their evolution over intermediate timescales. tRNA pools exhibit remarkable evolutionary churn in gene number, sequence, and isotype identity across evolutionary lineages (Rogers et al. 2010; Wang and Ruvinsky 2012; Yona et al. 2013; Rogers and Griffiths-Jones 2014; Velandia-Huerto et al. 2016) evolutionary patterns and the processes that drive them remain relatively underexplored, in part due to the same obstacles that impede functional analysis of cytosolic tRNAs: their large number in nuclear genomes and their complex organization. Gene orthology is a challenge to resolve syntenically across lineages, due to sequence similarity between loci as well as genuine gene gains and losses (Wang and Ruvinsky 2012; Velandia-Huerto et al. 2016; Lant et al. 2019). However, intraspecific variation in tRNA repertoires offers an entry point to elucidating how these gene sets diverge over time, by revealing early-stage changes that may include gene duplications, losses, or conversions. These analyses may also facilitate the discovery of TAM mechanisms specifically associated with Pol III transcription; tRNAs exhibit both heterogeneity and excess

mutation rates beyond those explained by known determinants (Seplyarskiy et al. 2023). Relatedly, mutational variation at tRNA genes can be leveraged to investigate codon usage and potential evolution within and across populations.

The most striking feature of tRNAs may be their conservation across the tree of life, and this universality appears to extend beyond the individual molecule to the operation of tRNAs as repertoires. The nuclear genomes of multicellular eukaryotes encode hundreds of cytosolic tRNAs characterized by complex multidimensionality (Goodenbour and Pan 2006; Bermudez-Santana et al. 2010). For example, the human genome encodes over 400 putatively functional cytosolic tRNAs that exhibit both shared and diverged sequence identities and diversity in the number of genes within each isoacceptor class (Goodenbour and Pan 2006; Chan and Lowe 2016). This complexity has been hypothesized to have evolved to meet the proteomic demands of humans as highly differentiated multicellular organisms (Pinzaru and Tavazoie 2023). Indeed, more complex organisms, like humans and chimpanzees, harbor more isodecoders than those that are less complex, like flies and worms (Goodenbour and Pan 2006; Orellana et al. 2022). Yet, this modest trend scales poorly with complexity. At only ~1000 cells (Sulston and Horvitz 1977), *C. elegans* is approximately 30 billion times smaller than a human but encodes over 500 tRNA genes and exhibits similar multidimensionality in distribution of isoacceptors (Duret 2000). While tRNA repertoires do exhibit lineage-specific diversification (McDonald et al. 2025), their general similarities suggest that tRNA functionality and the evolutionary pressures shaping tRNA pools are broadly conserved across eukaryotes and that discoveries in any system will be widely informative.

Comments on This Study

A central phenomenon in tRNA evolution is TAM, to which tRNA genes are highly exposed (Thornlow et al. 2018; Seplyarskiy et al. 2023). Here, we replicate initial reports in other eukaryotes to show evidence of elevated TAM at tRNA loci in *C. elegans*, *C. tropicalis*, and *C. briggsae* (Fig. 1c), further emphasizing the universality of this phenomenon (Thornlow et al. 2018). These findings complement our related work, in which we used a novel sequence substitution model to assess mutational variation in *C. elegans*, including its effects on molecular secondary structure and thermostability (Banos et al. 2025). Because tRNA genes are transcribed in all cells, TAM presumably occurs organism-wide, consistent with the high somatic mutation burden in tRNA genes in human cancer and aging (Murillo-Recio et al. 2025). However, only mutations arising in germline cells

contribute to the heritable variation observed within and across populations. This reasoning suggests that TAM is universal despite diverse mechanisms of eukaryotic germline formation, reinforcing predictions that population-level variation in tRNA genes imposes a mutation load on genomes across taxa.

The tRNA allelic variation we report here may compromise tRNA function by affecting folding and structure or amino acid proofreading or identity (Fig. 2). Our analysis indicates that strains differ in the number of functional tRNA genes. However, 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). Complete resolution of cytosolic tRNA complements for individual strains requires strain-specific assemblies and must distinguish true genomic differences from assembly artifacts that can arise even between assemblies of the same strain or organism (Murillo-Recio et al. 2025).

Isotype switching is a particularly compelling aspect of tRNA variation, considering its potential to compromise the protein sequence. We observed two broad categories of isotype switching (Fig. 3). First, we detected isotype mismatches that were fixed in the population, which likely represent ancestral changes and contribute to the isotype remodeling 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. These switches may predate speciation and may have fixed neutrally or adaptively, consistent with observations of leucine tRNAs in nematodes, humans, and other taxa (Rogers and Griffiths-Jones 2014; Lant et al. 2019); confirmation requires resolving orthology and synteny. Because the aminoacyl-tRNA synthetase for leucine, 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). We also observed fixed isotype switches from alanine and serine backbones. Second, we detected polymorphic isotype switches, present in a subset of strains, in all three species. These variants were rare, occurred at low frequency, and based on their tRNA backbones, are unlikely to 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.

The standing variation at tRNA genes that we report here may persist longer in these three nematode species because of their androdioecious mating system; extended selfing with low rates of outcrossing promotes the accumulation of mutational load (Loewe and Cutter 2008). Background selection dominates in these genomes, but varies between the low-recombination centers of chromosomes and the higher-recombination arms (Rockman et al. 2010; Andersen et al. 2012). We show that tRNAs are disproportionately distributed on arms (Fig. 4), which may reflect an adaptation for more efficient purging of deleterious mutations. This distribution may also facilitate the rapid birth-death cycles that characterize tRNA gene families (Nei and Rooney 2005; Wang and Ruvinsky 2012), given the greater potential for duplication, excision, and conversion under higher crossover rates. 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 arm-biased (Surzycki and Belknap 2000; Laricchia et al. 2017), so the tRNA-specific distribution might reflect tRNA-specific evolution.

The genetic variation in tRNAs we report here may confer biological significance beyond their canonical role in translation; alternative activities of tRNAs continue to be discovered (reviewed in (Schimmel 2018)). Differences in tRNA genes may affect tRNA participation in protein complexes and the generation of tRNA fragments; variation in human isodecoder gene sets has been observed to impact expression levels of tRNA fragments, which have diverse roles beyond those of full-length tRNAs (Torres 2019; Torres et al. 2019). As previously mentioned, the role of genetic variation on tRNA expression and regulation, including via posttranscriptional modifications, remains an outsized and essential question.

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 622 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.

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. We ran tRNAscan-SE using default settings, including full logs, progress reports, and isotype-specific modeling by including options *-log*, *-progress*, *-detail*, and *-isospecific*. The tRNAscan-SE output from the reference genomes thus established the set of tRNA genes (and pseudogenes) in our study; it also provided, for each gene: 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. We also sought to maximize inclusivity when establishing the initial gene sets from the reference genomes, given our interest in mutational variation in tRNAs. 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). Following downstream analyses in which we interrogated all alleles for functional status, we were also able to classify genes according to whether none, some, or all their alleles were predicted to be nonfunctional.

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 recorded the identity of each variant that occurred within 40 bp 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 androdieious 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 the directory *flank_specific*.

Allele Frequency Analyses of Variants in tRNAs 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 (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 the 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 nonmissing 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 for multiple site subsets as missingness impacts the calculation: sites with $>80\%$ or $>90\%$ samples with non-missing genotypes. We also performed the calculation 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 the 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-like sequences and surrounding regions (identified by tRNAscan-SE as described above), then generated allele-specific FASTA files by inserting each observed variant(s) for each strain, overwriting the reference base at that position (SNVs and INDELS included). Strains with missing genotypes at any variant within a tRNA were not assigned an allele identity for that gene and were instead recorded as missing. To characterize the alleles, we ran tRNAscan-SE on the FASTA file containing all tRNA allele sequences (options as used for 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 "non-functional" 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/non-functional 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 we 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% to 61% of variable (low-frequency) isotype switches were attributed to anticodon mutations and 39% to 44% to backbone mutations. In all backbone mutation cases and 58% to 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 we inferred that 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 pseudogenes, we could not map all bases in the tRNA to a predicted secondary structure, so we flagged these 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 versus 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 relevant 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, and 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 (i) for determining the sequence alignment distance for all pairs of unique sequences and (ii) 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) (Mike and Davis 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), *circclize* (v0.4.16) (Gu et al. 2014), *magick* (v2.8.7) (Ooms 2025a), and *pdftools* (v3.5.0) (Ooms 2025b).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

The authors thank members of the Southeast Center for Mathematics and Biology, Abigail Lind and members of the Lind lab, Rohini Janivara, and Abbe LaBella 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 the NSF-Simons Southeast Center for Mathematics and Biology (SCMB) through the grants National Science Foundation (NSF) DMS-1764406 and Simons Foundation/SFARI 594594, by National Science Foundation EDGE grant 2319796 (to ABP), and by support from the Georgia Institute of Technology.

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 are interactively available via web app at <https://wild.worm.biosci.gatech.edu/trnarefalleles/>. Code used in this study is available at <https://github.com/paabylab/trnavcfalleles>.

Literature Cited

- Ahlmann-Eltze C, Patil I. 2021. ggsignif: R Package for Displaying Significance Brackets for 'ggplot2' [preprint]. PsyArxiv. <https://doi.org/10.31234/osf.io/7awm6>.
- Andersen EC, et al. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet.* 2012;44:285–290. <https://doi.org/10.1038/ng.1050>.
- Andersen EC, Rockman MV. Natural genetic variation as a tool for discovery in *Caenorhabditis* nematodes. *Genetics.* 2022;220:iyab156. <https://doi.org/10.1093/genetics/iyab156>.
- Aphalo PJ. ggpmisc: Miscellaneous Extensions to 'ggplot2'. 2024. <https://CRAN.R-project.org/package=ggpmisc>
- Banos H, Wang L, Simonti C, Paaby A, Heitsch C. How does transcription-associated mutagenesis shape tRNA microevolution? *GBE.* 2025. <https://doi.org/10.1093/gbe/evag050>.
- Barnes TM, Kohara Y, Coulson A, Hekimi S. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics.* 1995;141:159–179. <https://www.ncbi.nlm.nih.gov/pubmed/8536965> <https://doi.org/10.1093/genetics/141.1.159>.
- Barriere A, Felix MA. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. *Curr Biol.* 2005a;15:1176–1184. <https://doi.org/10.1016/j.cub.2005.06.022>.
- Barriere A, Felix MA. Natural variation and population genetics of *Caenorhabditis elegans*. *WormBook.* 2005b;1:1–19. <https://doi.org/10.1895/wormbook.1.43.1>.
- Berg MD, et al. Targeted sequencing reveals expanded genetic diversity of human transfer RNAs. *RNA Biol.* 2019;16:1574–1585. <https://doi.org/10.1080/15476286.2019.1646079>.
- Bermudez-Santana C, et al. Genomic organization of eukaryotic tRNAs. *BMC Genomics.* 2010;11:270. <https://doi.org/10.1186/1471-2164-11-270>.
- Bernstein MR, Rockman MV. Fine-scale crossover rate variation on the *Caenorhabditis elegans* X chromosome. *G3 (Bethesda).* 2016;6:1767–1776. <https://doi.org/10.1534/g3.116.028001>.
- Bida JP, Maher LJ, 3rd. Improved prediction of RNA tertiary structure with insights into native state dynamics. *RNA.* 2012;18(3):385–393. <https://doi.org/10.1261/rna.027201.111>.
- Bloom-Ackermann Z, et al. A comprehensive tRNA deletion library unravels the genetic architecture of the tRNA pool. *PLoS Genet.* 2014;10:e1004084. <https://doi.org/10.1371/journal.pgen.1004084>.
- Campitelli E. ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'. 2025. <https://CRAN.R-project.org/package=ggnewscale>
- Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 2021;49:9077–9096. <https://doi.org/10.1093/nar/gkab688>.
- Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 2009;37:D93–D97. <https://doi.org/10.1093/nar/gkn787>.
- Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* 2016;44:D184–D189. <https://doi.org/10.1093/nar/gkv1309>.
- Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol.* 2019;1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.
- Commans S, Bock A. Selenocysteine inserting tRNAs: an overview. *FEMS Microbiol Rev.* 1999;23:335–351. <https://doi.org/10.1111/j.1574-6976.1999.tb00403.x>.
- Crick FH. Codon–anticodon pairing: the wobble hypothesis. *J Mol Biol.* 1966;19:548–555. [https://doi.org/10.1016/s0022-2836\(66\)80022-0](https://doi.org/10.1016/s0022-2836(66)80022-0).
- Crombie TA, et al. Deep sampling of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations. *Elife.* 2019;8:e50465. <https://doi.org/10.7554/eLife.50465>.
- Crombie TA, et al. CaenDR, the *Caenorhabditis* natural diversity resource. *Nucleic Acids Res.* 2024;52:D850–D858. <https://doi.org/10.1093/nar/gkad887>.
- Danecek P, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10:giab008. <https://doi.org/10.1093/giga-science/giab008>.
- Darrow EM, Chadwick BP. 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.* 2014;42:6421–6435. <https://doi.org/10.1093/nar/gku280>.
- Davey-Young J, et al. Mistranslating the genetic code with leucine in yeast and mammalian cells. *RNA Biol.* 2024;21:507–529. <https://doi.org/10.1080/15476286.2024.2340297>.
- Dolgin ES, Charlesworth B, Baird SE, Cutter AD. Inbreeding and outbreeding depression in *Caenorhabditis* nematodes. *Evolution.* 2007;61:1339–1352. <https://doi.org/10.1111/j.1558-5646.2007.00118.x>.
- Dowle M, Srinivasan A. data.table: Extension of 'data.frame'. 2022. <https://Rdatatable.gitlab.io/data.table>
- Duret L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* 2000;16:287–289. [https://doi.org/10.1016/s0168-9525\(00\)02041-2](https://doi.org/10.1016/s0168-9525(00)02041-2).
- Félix M-A, et al. Species richness, distribution and genetic diversity of *Caenorhabditis* nematodes in a remote tropical rainforest.

- BMC Evolutionary Biology. 2013;13:10. <https://doi.org/10.1186/1471-2148-13-10>.
- Florentz C, Sohm B, Tryoen-Toth P, Putz J, Sissler M. Human mitochondrial tRNAs in health and disease. *Cell Mol Life Sci*. 2003;60:1356–1375. <https://doi.org/10.1007/s00018-003-2343-1>.
- Frézal L, Félix M-A. *C. elegans* outside the Petri dish. *eLife*. 2015;4:e05849. <https://doi.org/10.7554/eLife.05849>.
- Fusca DD, Dall'Acqua MN, Sánchez-Ramírez S, Cutter AD. Phylogenomic timetree-calibrated speciation clocks for *Caenorhabditis* nematodes reveal slow but disproportionate accumulation of post-zygotic reproductive isolation. *PLOS Genetics*. 2025;21:e1011852. <https://doi.org/10.1371/journal.pgen.1011852>.
- Gaillard H, Aguilera A. Transcription as a threat to genome integrity. *Annu Rev Biochem*. 2016;85:291–317. <https://doi.org/10.1146/annurev-biochem-060815-014908>.
- Galli G, Hofstetter H, Birnstiel ML. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature*. 1981;294:626–631. <https://doi.org/10.1038/294626a0>.
- Gao L, et al. Selective gene expression maintains human tRNA anticodon pools during differentiation. *Nat Cell Biol*. 2024;26:100–112. <https://doi.org/10.1038/s41556-023-01317-3>.
- Giege R, Eriani G. The tRNA identity landscape for aminoacylation and beyond. *Nucleic Acids Res*. 2023;51:1528–1570. <https://doi.org/10.1093/nar/gkad007>.
- Giege R, Sissler M, Florentz C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res*. 1998;26:5017–5035. <https://doi.org/10.1093/nar/26.22.5017>.
- Goodenbour JM, Pan T. Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res*. 2006;34:6137–6146. <https://doi.org/10.1093/nar/gkl725>.
- Goto Y, Nonaka I, Horai S. A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature*. 1990;348:651–653. <https://doi.org/10.1038/348651a0>.
- Green P, et al. Transcription-associated mutational asymmetry in mammalian evolution. *Nat Genet*. 2003;33:514–517. <https://doi.org/10.1038/ng1103>.
- Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30:2811–2812. <https://doi.org/10.1093/bioinformatics/btu393>.
- Harris CR, et al. Array programming with NumPy. *Nature*. 2020;585:357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
- Hillier LW, et al. Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. *PLoS Biol*. 2007;5:e167. <https://doi.org/10.1371/journal.pbio.0050167>.
- Hipps D, Shiba K, Henderson B, Schimmel P. Operational RNA code for amino acids: species-specific aminoacylation of minihelices switched by a single nucleotide. *Proc Natl Acad Sci U S A*. 1995;92:5550–5552. <https://doi.org/10.1073/pnas.92.12.5550>.
- Hughes LA, et al. Copy number variation in tRNA isodecoder genes impairs mammalian development and balanced translation. *Nat Commun*. 2023;14:2210. <https://doi.org/10.1038/s41467-023-37843-9>.
- Iben JR, Maraia RJ. tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA*. 2012;18:1358–1372. <https://doi.org/10.1261/rna.032151.111>.
- Iben JR, Maraia RJ. tRNA gene copy number variation in humans. *Gene*. 2014;536:376–384. <https://doi.org/10.1016/j.gene.2013.11.049>.
- Jinks-Robertson S, Bhagwat AS. Transcription-associated mutagenesis. *Annu Rev Genet*. 2014;48:341–359. <https://doi.org/10.1146/annurev-genet-120213-092015>.
- Kaur T, Rockman MV. Crossover heterogeneity in the absence of hotspots in *Caenorhabditis elegans*. *Genetics*. 2014;196:137–148. <https://doi.org/10.1534/genetics.113.158857>.
- Kobayashi Y, et al. 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*. 1990;173:816–822. [https://doi.org/10.1016/s0006-291x\(05\)80860-5](https://doi.org/10.1016/s0006-291x(05)80860-5).
- Kutter C, et al. Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nat Genet*. 2011;43:948–955. <https://doi.org/10.1038/ng.906>.
- Lant JT, Berg MD, Heinemann IU, Brandl CJ, O'Donoghue P. Pathways to disease from natural variations in human cytoplasmic tRNAs. *J Biol Chem*. 2019;294:5294–5308. <https://doi.org/10.1074/jbc.REV118.002982>.
- Laricchia K, Zdravljec S, Cook D, Andersen E. Natural variation in the distribution and abundance of transposable elements across the *Caenorhabditis elegans* species. *Molecular biology and evolution*. 2017;34:2187–2202. <https://doi.org/10.1093/molbev/msx155>.
- Larkin MA, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23:2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>.
- Lee D, et al. Balancing selection maintains hyper-divergent haplotypes in *Caenorhabditis elegans*. *Nat Ecol Evol*. 2021;5:794–807. <https://doi.org/10.1038/s41559-021-01435-x>.
- Leung J, Sinclair DA, Hayashi S, Tener GM, Grigliatti TA. Informational redundancy of tRNA(4Ser) and tRNA(7Ser) genes in *Drosophila melanogaster* and evidence for intergenic recombination. *J Mol Biol*. 1991;219:175–188. [https://doi.org/10.1016/0022-2836\(91\)90560-s](https://doi.org/10.1016/0022-2836(91)90560-s).
- Li C, Qian W, Maclean CJ, Zhang J. The fitness landscape of a tRNA gene. *Science*. 2016;352:837–840. <https://doi.org/10.1126/science.aae0568>.
- Loewe L, Cutter AD. On the potential for extinction by Muller's ratchet in *Caenorhabditis elegans*. *BMC Evol Biol*. 2008;8:125. <https://doi.org/10.1186/1471-2148-8-125>.
- McDonald DW, Joos L, Duenwald ML. Implications of codon usage, tRNA gene redundancy and tRNA gene clustering in experimental models of mistranslation. *J Mol Biol*. 2025;438:169573. <https://doi.org/10.1016/j.jmb.2025.169573>.
- Mike FC, Davis TL. ggplot2 authors, ggpattern: 'ggplot2' Pattern Geoms. 2022. <https://CRAN.R-project.org/package=ggpattern>
- Moya ND, et al. 2025a Dec 11. *Caenorhabditis briggsae* ancestral genomic hyper-diversity contrasts with globally distributed genome-wide haplotypes [preprint]. *bioRxiv*:2025.2012.2008.693002. <https://doi.org/10.64898/2025.12.08.693002>.
- Moya ND, Yan SM, McCoy RC, Andersen EC. The long and short of hyperdivergent regions. *Trends Genet*. 2025b;41:303–314. <https://doi.org/10.1016/j.tig.2024.11.005>.
- Murillo-Recio M, et al. 2025 June 9. Characterization of somatic mutations in human tRNA genes reveals tumor-specific mutational loads, and the generation of tRNA variants that alter the genetic code [preprint]. *bioRxiv*:2025.2006.2009.658696. <https://doi.org/10.1101/2025.06.09.658696>.
- Murrell P. gridBase: Integration of base and grid graphics. 2014. <https://CRAN.R-project.org/package=gridBase>
- Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29:2933–2935. <https://doi.org/10.1093/bioinformatics/btt509>.

- Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet.* 2005;39:121–152. <https://doi.org/10.1146/annurev.genet.39.073003.112240>.
- Neuwirth E. RColorBrewer: ColorBrewer Palettes. 2022. <https://CRAN.R-project.org/package=RColorBrewer>
- Noble LM, et al. Selfing is the safest sex for *Caenorhabditis tropicalis*. *Elife.* 2021;10:e62587. <https://doi.org/10.7554/eLife.62587>.
- Ooms J. magick: Advanced Graphics and Image-Processing in R. 2025a. <https://doi.org/10.32614/CRAN.package.magick>.
- Ooms J. pdftools: Text Extraction, Rendering and Converting of PDF Documents. 2025b. <https://doi.org/10.32614/CRAN.package.pdfutils>.
- Orellana EA, Siegal E, Gregory RI. tRNA dysregulation and disease. *Nat Rev Genet.* 2022;23:651–664. <https://doi.org/10.1038/s41576-022-00501-9>.
- Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 2019;35:526–528. <https://doi.org/10.1093/bioinformatics/bty633>.
- Parisien M, Wang X, Pan T. Diversity of human tRNA genes from the 1000-genomes project. *RNA Biol.* 2013;10:1853–1867. <https://doi.org/10.4161/rna.27361>.
- Pedersen TL. ggforce: Accelerating 'ggplot2'. 2022. <https://CRAN.R-project.org/package=ggforce>
- Perona JJ, Gruic-Sovulj I. Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. In: Kim S, editors. *Aminoacyl-tRNA synthetases in biology and medicine*. Springer Netherlands; 2014. p. 1–41.
- Pinzaru AM, Tavazoie SF. Transfer RNAs as dynamic and critical regulators of cancer progression. *Nat Rev Cancer.* 2023;23:746–761. <https://doi.org/10.1038/s41568-023-00611-4>.
- Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* 2012;8:e1002603. <https://doi.org/10.1371/journal.pgen.1002603>.
- R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2025. <https://www.R-project.org/>
- Ren K, Russell K. formattable: create 'Formattable' Data Structures. 2021. <https://CRAN.R-project.org/package=formattable>
- Rockman MV, Kruglyak L. Recombinational landscape and population genetics of *Caenorhabditis elegans*. *PLOS Genetics.* 2009;5:e1000419. <https://doi.org/10.1371/journal.pgen.1000419>.
- Rockman MV, Skrovaneck SS, Kruglyak L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science.* 2010;330:372–376. <https://doi.org/10.1126/science.1194208>.
- Rogers HH, Bergman CM, Griffiths-Jones S. The evolution of tRNA genes in *Drosophila*. *Genome Biol Evol.* 2010;2:467–477. <https://doi.org/10.1093/gbe/evq034>.
- Rogers HH, Griffiths-Jones S. tRNA anticodon shifts in eukaryotic genomes. *RNA.* 2014;20:269–281. <https://doi.org/10.1261/rna.041681.113>.
- Ross JA, et al. *Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. *PLoS Genet.* 2011;7:e1002174. <https://doi.org/10.1371/journal.pgen.1002174>.
- Roth AC. Decoding properties of tRNA leave a detectable signal in codon usage bias. *Bioinformatics.* 2012;28:i340–i348. <https://doi.org/10.1093/bioinformatics/bts403>.
- Schimmel P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol.* 2018;19:45–58. <https://doi.org/10.1038/nrm.2017.77>.
- Schliep KP. Phangorn: phylogenetic analysis in R. *Bioinformatics.* 2011;27:592–593. <https://doi.org/10.1093/bioinformatics/btq706>.
- Schoenmakers E, et al. Mutation in human selenocysteine transfer RNA selectively disrupts selenoprotein synthesis. *J Clin Invest.* 2016;126:992–996. <https://doi.org/10.1172/JCI84747>.
- Septyarskiy V, et al. A mutation rate model at the basepair resolution identifies the mutagenic effect of polymerase III transcription. *Nat Genet.* 2023;55:2235–2242. <https://www.ncbi.nlm.nih.gov/pubmed/38036792>. <https://doi.org/10.1038/s41588-023-01562-0>.
- Shih DJH. argparser: Command-Line Argument Parser. 2021. <https://CRAN.R-project.org/package=argparser>.
- Sternberg PW, et al. WormBase 2024: status and transitioning to Alliance infrastructure. *Genetics.* 2024;227:iyae050. <https://www.ncbi.nlm.nih.gov/pubmed/38573366>. <https://doi.org/10.1093/genetics/iyae050>.
- Stevens L, et al. Chromosome-level reference genomes for two strains of *Caenorhabditis briggsae*: an improved platform for comparative genomics. *Genome Biol Evol.* 2022;14:evac042. <https://www.ncbi.nlm.nih.gov/pubmed/35348662>. <https://doi.org/10.1093/gbe/evac042>.
- Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol.* 1977;56:110–156. <https://www.ncbi.nlm.nih.gov/pubmed/838129>. [https://doi.org/10.1016/0012-1606\(77\)90158-0](https://doi.org/10.1016/0012-1606(77)90158-0).
- Surzycki SA, Belknap WR. Repetitive-DNA elements are similarly distributed on *Caenorhabditis elegans* autosomes. *Proc Natl Acad Sci U S A.* 2000;97:245–249. <https://doi.org/10.1073/pnas.97.1.245>.
- Suzuki T. The expanding world of tRNA modifications and their disease relevance. *Nat Rev Mol Cell Biol.* 2021;22:375–392. <https://www.ncbi.nlm.nih.gov/pubmed/33658722>. <https://doi.org/10.1038/s41580-021-00342-0>.
- Suzuki T, Nagao A, Suzuki T. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. *Annu Rev Genet.* 2011;45:299–329. <https://www.ncbi.nlm.nih.gov/pubmed/21910628>. <https://doi.org/10.1146/annurev-genet-110410-132531>.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123:585–595. <https://www.ncbi.nlm.nih.gov/pubmed/2513255>. <https://doi.org/10.1093/genetics/123.3.585>.
- Tennakoon R, Bily TMI, Hasan F, Hoffman KS, O'Donoghue P. Natural human tRNAAla anticodon variants mistranslate the genetic code. *RNA.* 2025;31:791–806. <http://rnajournal.cshlp.org/content/31/6/791.abstract>. <https://doi.org/10.1261/rna.080450.125>.
- Thornlow BP, et al. Transfer RNA genes experience exceptionally elevated mutation rates. *Proc Natl Acad Sci U S A.* 2018;115:8996–9001. <https://www.ncbi.nlm.nih.gov/pubmed/30127029>. <https://doi.org/10.1073/pnas.1801240115>.
- Thornlow BP, et al. Predicting transfer RNA gene activity from sequence and genome context. *Genome Res.* 2020;30:85–94. <https://www.ncbi.nlm.nih.gov/pubmed/31857444>. <https://doi.org/10.1101/gr.256164.119>.
- Torres AG. Enjoy the silence: nearly half of human tRNA genes are silent. *Bioinform Biol Insights.* 2019;13:1177932219868454. <https://www.ncbi.nlm.nih.gov/pubmed/31447549>. <https://doi.org/10.1177/1177932219868454>.
- Torres AG, Reina O, Stephan-Otto Attolini C, Ribas de Pouplana L. Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. *Proc Natl Acad Sci U S A.* 2019;116:8451–8456. <https://www.ncbi.nlm.nih.gov/pubmed/30962382>. <https://doi.org/10.1073/pnas.1821120116>.

- van den Brand T. ggh4x: Hacks for 'ggplot2'. 2025. <https://CRAN.R-project.org/package=ggh4x>.
- Velandia-Huerto CA, et al. Orthologs, turn-over, and remodeling of tRNAs in primates and fruit flies. *BMC Genomics*. 2016;17:617. <https://www.ncbi.nlm.nih.gov/pubmed/27515907>. <https://doi.org/10.1186/s12864-016-2927-4>.
- Wang PP, Ruvinsky I. Family size and turnover rates among several classes of small non-protein-coding RNA genes in *Caenorhabditis* nematodes. *Genome Biol Evol*. 2012;4:565–574. <https://www.ncbi.nlm.nih.gov/pubmed/22467905>. <https://doi.org/10.1093/gbe/evs034>.
- Wickham H. *Ggplot2: elegant graphics for data analysis*. Springer-Verlag New York; 2016. <https://ggplot2.tidyverse.org>
- Wickham H, Pedersen T, Seidel D. scales: Scale Functions for Visualization. 2025. <https://CRAN.R-project.org/package=scales>
- Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. 2020. <https://CRAN.R-project.org/package=cowplot>.
- Wolf M, Koetschan C, Muller T. ITS2, 18S, 16S or any other RNA—simply aligning sequences and their individual secondary structures simultaneously by an automatic approach. *Gene*. 2014;546:145–149. <https://www.ncbi.nlm.nih.gov/pubmed/24881812>. <https://doi.org/10.1016/j.gene.2014.05.065>.
- Yang J. Tajima's D and genomic features. 2022. <https://jyanglab.com/slides/2022-agro932/w6lab.html#2>.
- Yarham JW, Elson JL, Blakely EL, McFarland R, Taylor RW. Mitochondrial tRNA mutations and disease. *Wiley Interdiscip Rev RNA*. 2010;1:304–324. <https://www.ncbi.nlm.nih.gov/pubmed/21935892>. <https://doi.org/10.1002/wrna.27>.
- Yona AH, et al. tRNA genes rapidly change in evolution to meet novel translational demands. *Elife*. 2013;2:e01339. <https://www.ncbi.nlm.nih.gov/pubmed/24363105>. <https://doi.org/10.7554/eLife.01339>.
- Yu G. scatterpie: Scatter Pie Plot. 2025. <https://CRAN.R-project.org/package=scatterpie>.
- Yu G. Data integration, manipulation and visualization of phylogenetic trees. Chapman and Hall/CRC; 2022. p. 276. <https://www.taylorfrancis.com/books/mono/10.1201/9781003279242/data-integration-manipulation-visualization-phylogenetic-trees-guangchuang-yu>. <https://doi.org/10.1201/9781003279242>.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. Ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*. 2017;8:28–36. <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12628>. <https://doi.org/10.1111/2041-210X.12628>.

Associate editor: Charles Baer

Supplemental Figures

Allelic variation at tRNA genes in three nematode species indicates mutation load despite strong purifying selection

Avery Davis Bell¹, Corinne Simonti^{1,3}, Hector Baños^{2,4}, Ling Wang¹, Christine Heitsch², Joseph Lachance¹, 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

³Current address: AAAS, Washington, DC

⁴Current address: Department of Mathematics, California State University San Bernardino, San Bernardino, CA, USA

*Correspondence: paaby@gatech.edu

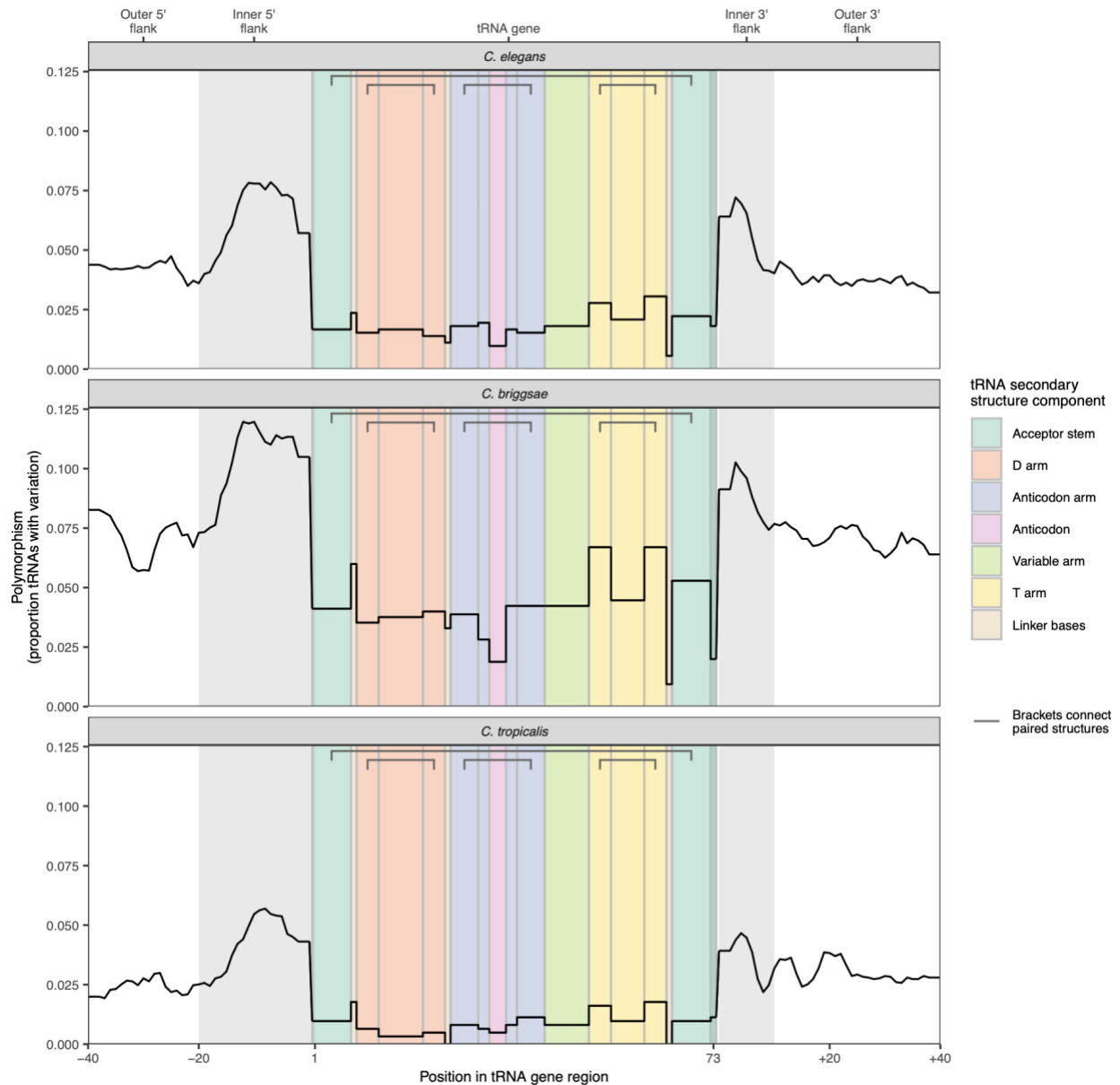


Figure S1. High rate of mutation in tRNA flanking regions in *Caenorhabditis* nematodes. As in **Figure 1C** but for all three species (sub-plot titles). Linker bases are included here (omitted from Figure 1); these do not occur in all tRNAs and therefore estimates for them are unreliable.

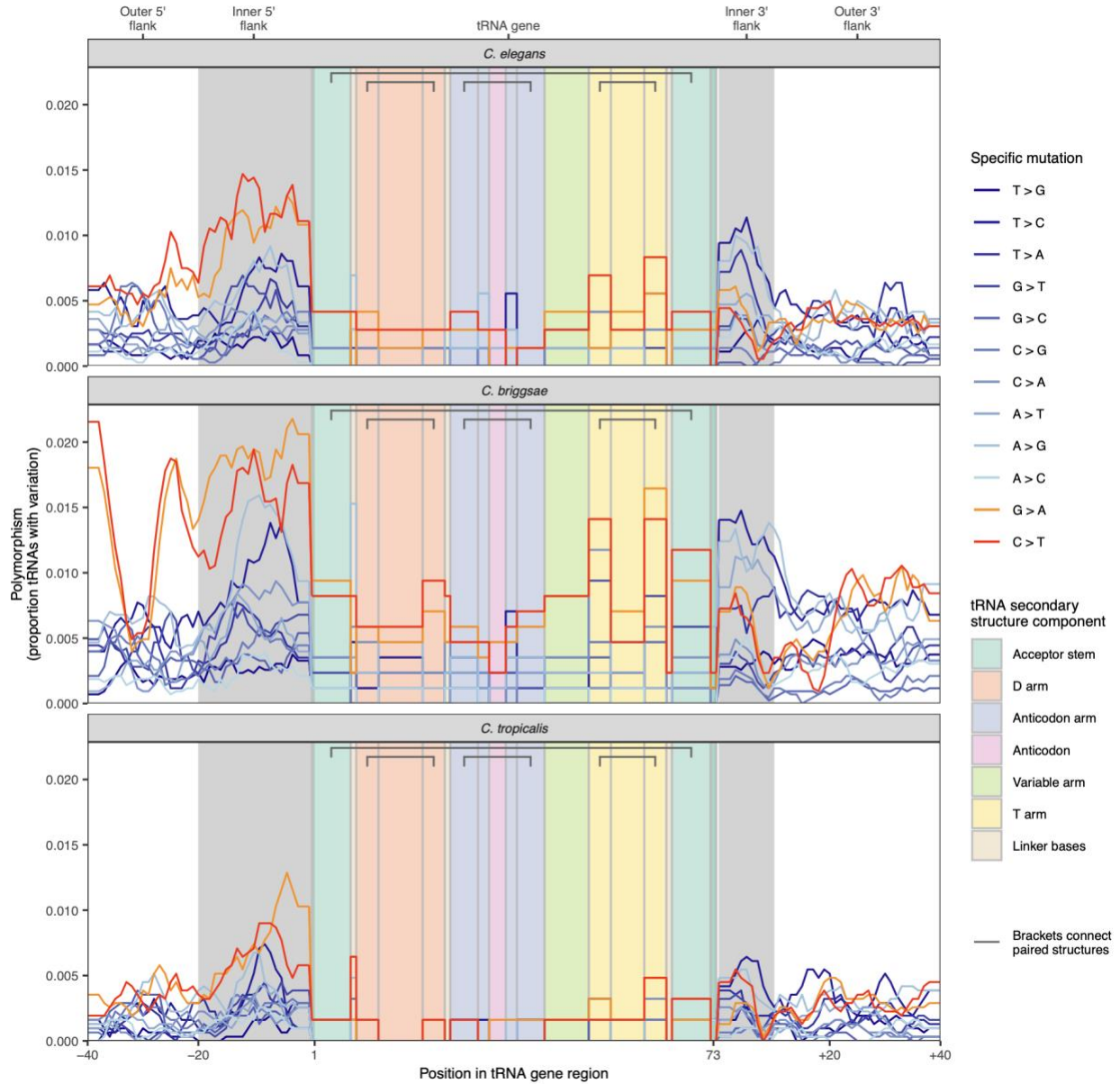


Figure S2. Signature of transcription-associated mutagenesis in *Caenorhabditis*. As in **Figure 1C**, but each class of SNV mutation is plotted separately in its own color, with TAM-characterizing mutations in orange and red. Linker bases are included here (omitted from Figure 1); these do not occur in all tRNAs and therefore estimates for them are unreliable.

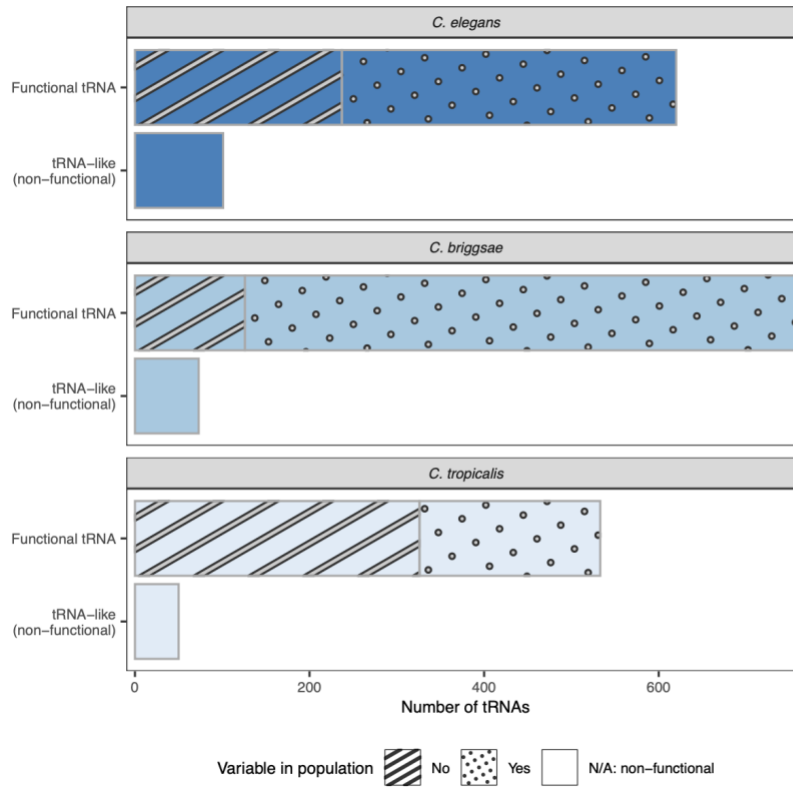


Figure S3. Total numbers of tRNA genes and tRNA-like sequences (pseudo genes) across *Caenorhabditis* species, subdivided by the number of genes with multiple alleles in the population (variable) vs fixed in the population (not variable).

(Related to **Figure 2**)

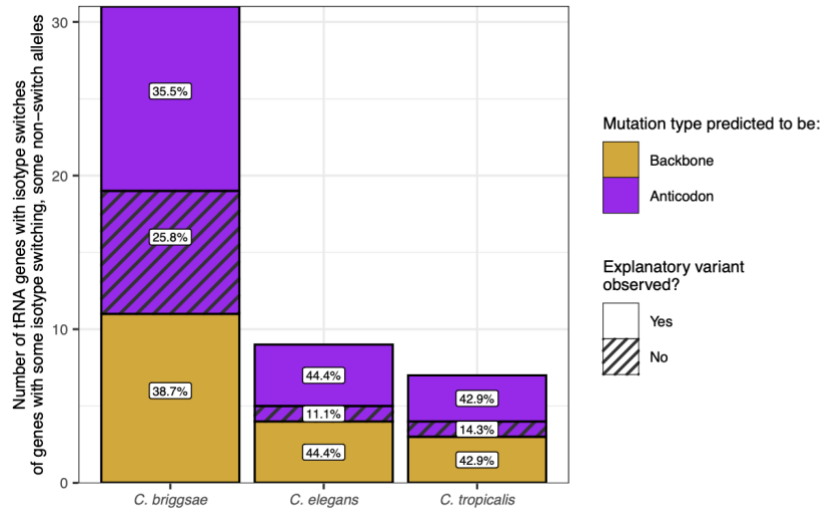


Figure S4. Characterization of variable isotype switch alleles by whether we can independently from tRNAscan-SE identify mutations likely underlying the isotype switch from the genetic variants. The proportion of each bar (per species) is shown on the plot, with the top and bottom bars summing to the total proportion of tRNAscan-SE called isotype switches for which we see independent mutational evidence. This is only for the ‘some switches, some not’ category (purple in **Figure 3A**) because if the isotype switch is fixed there is no variation in the population hence we cannot observe a variant causing the isotype switch. An explanatory variant not being observed could occur due to several factors including that the variant was an INDEL, which our methods cannot accurately map to structures (so they are included in allelic diversity but not per-variant analyses) and that the tRNA gene sequence was not appropriately broken down into secondary structure components and so we cannot analyze specific variants for that tRNA.

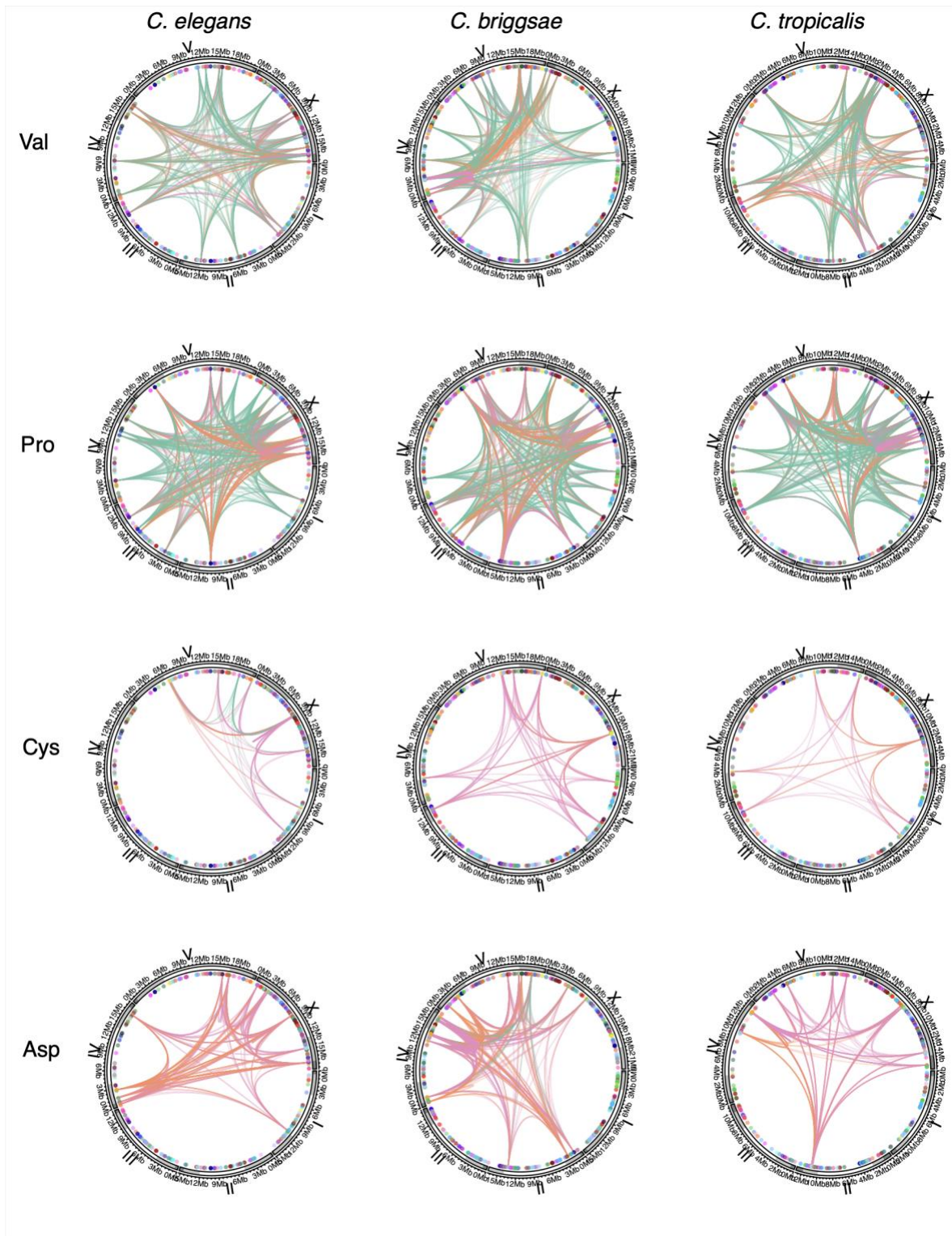


Figure S5. Distribution of tRNA genes in reference genomes and connections between tRNAs of specific isotypes. As in **Figure 4B** (including colors, see that key). All tRNAs are shown as dots in all plots in their genomic position (circular layout) per species (columns). The connections between all tRNAs with a given isotype (rows) are plotted.

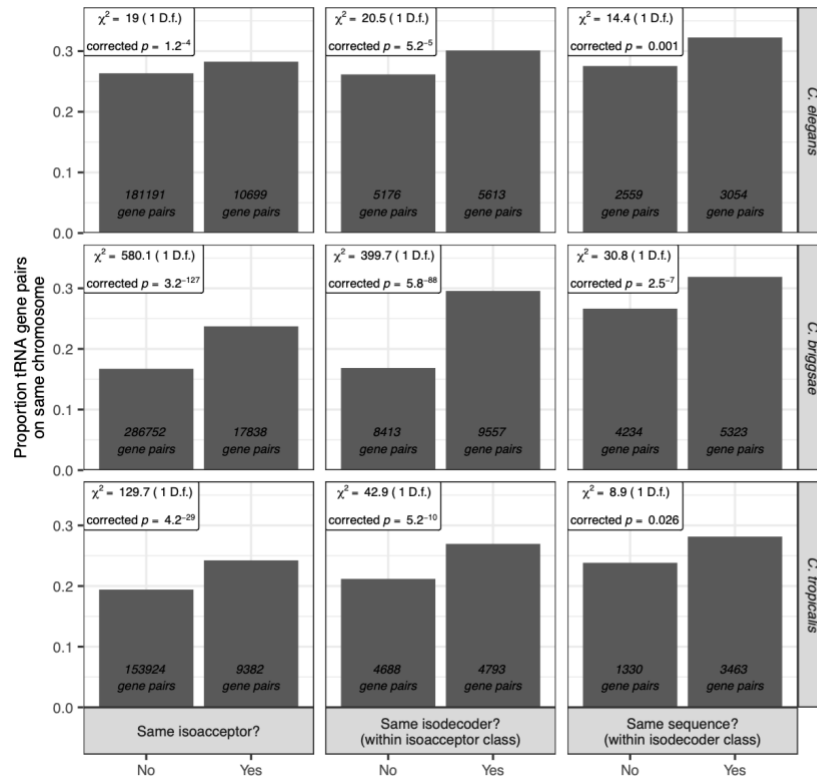


Figure S6. The proportion of tRNA gene pairs occurring on the same chromosome in the reference genome in three species (rows), subdivided by gene similarity or difference (columns). Statistical test results are shown for each comparison in each species; p-values shown are two-sided and Bonferroni corrected for multiple tests ($n = 9$ tests; 3 species x 3 tests). All tRNA gene pairs between genes with at least one non-functional allele are included; n pairs annotated on bars.

(Related to **Figure 4**)

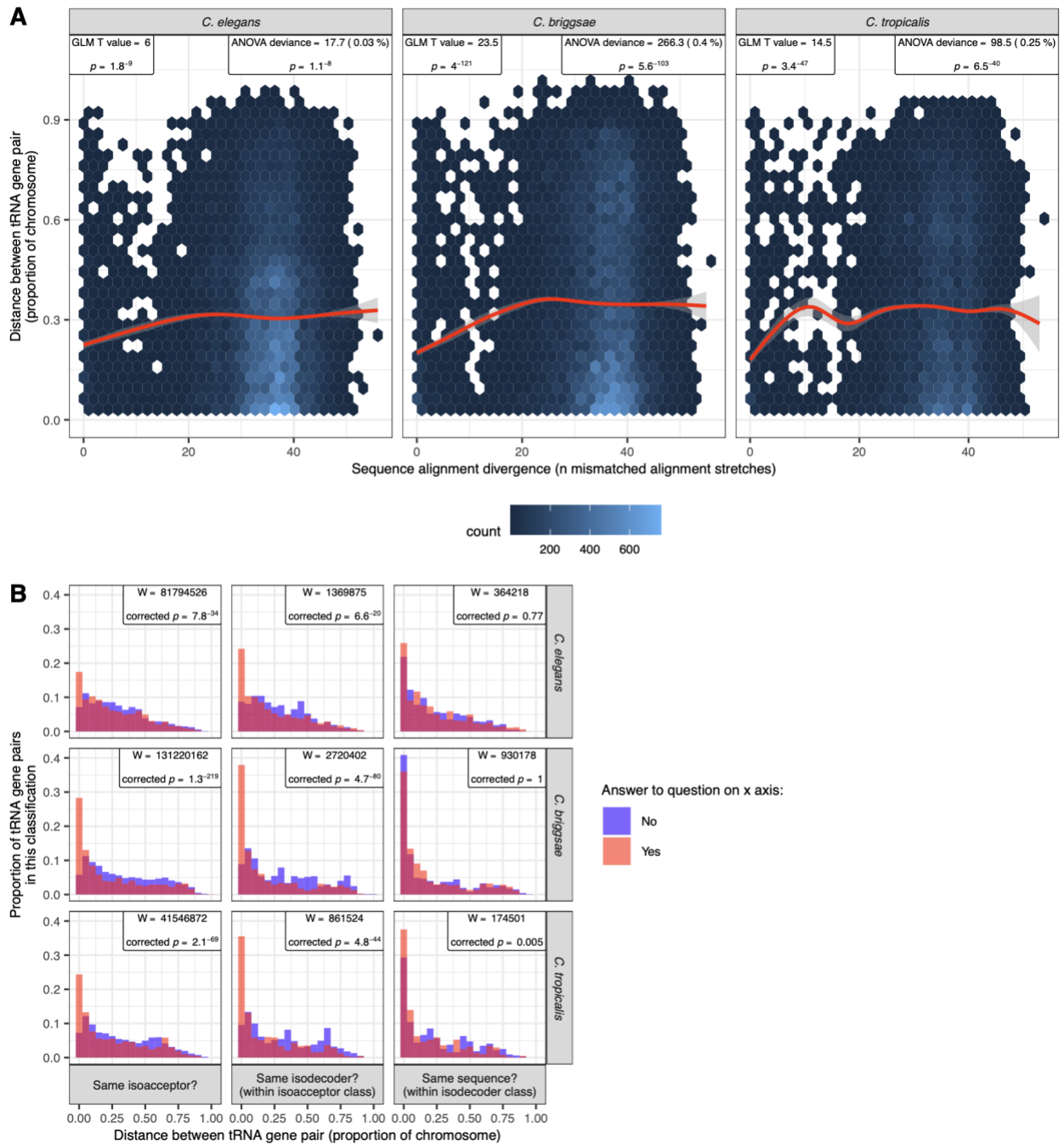


Figure S7. Relationships between tRNA gene sequence similarity and distance separating tRNA gene pairs in the genome. **A**, Quantitative relationship between sequence alignment divergence between tRNA pairs (x axis, see Methods) and distance between tRNA pairs (y axis), for all gene pairs occurring on the same chromosome. Statistical test results are shown from a generalized linear model (gamma log) modeling distance between genes as a function of chromosome, pseudogene status, and sequence divergence; ANOVA results show the deviance and percent deviance explained by sequence divergence from this GLM and GLM results are for the sequence divergence term. **B**, Distribution of distance between tRNA gene pairs when the pair does or does not share the provided characteristic (overlaid histograms; height of bar

corresponds to proportion of the pairs in that category in that species). Statistical test results for Mann-Whitney tests are shown; p-values are two-sided and Bonferroni-corrected for multiple hypothesis testing ($\times 9$: number of species \times number of tests). Genes with all alleles classified as nonfunctional are excluded from this analysis (but included in A where pseudogene status can be included in statistical modeling).

(Related to **Figure 4**)

Supplemental Tables

Allelic variation at tRNA genes in three nematode species indicates mutation load despite strong purifying selection

Avery Davis Bell¹, Corinne Simonti^{1,3}, Hector Baños^{2,4}, Ling Wang¹, Christine Heitsch², Joseph Lachance¹, 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

³Current address: AAAS, Washington, DC

⁴Current address: Department of Mathematics, California State University San Bernardino, San Bernardino, CA, USA

*Correspondence: paaby@gatech.edu

Table S1. Number of tRNAs encoding each amino acid in three *Caenorhabditis* species. Each number is the number of genes of that predicted backbone (tRNAScan-SE AlleleCM) in the reference genome, excluding genes where the reference allele was predicted to be nonfunctional.

Amino acid	Number of tRNA genes		
	<i>C. elegans</i>	<i>C. briggsae</i>	<i>C. tropicalis</i>
Ala	37	48	36
Arg	45	48	36
Asn	20	22	16
Asp	28	35	24
Cys	14	16	10
SeC	1	2	1
Gln	27	36	27
Glu	41	54	38
Gly	51	80	50
His	20	18	14
Ile	24	29	23
Leu	50	78	54
Lys	46	54	37
Met	9	13	11
iMet	9	10	9
Phe	15	21	14
Pro	42	57	51
Ser	40	41	31
Thr	33	40	32
Trp	12	16	13
Tyr	19	21	12
Val	32	35	31

Table S2. Number of tRNA genes with each codon (DNA) in three *Caenorhabditis* species. Codons are sorted by the amino acid each codon matches (second column; canonical amino acid this codon matches ignoring possible isotype switching). The number is for all genes in the reference genome, excluding genes where the reference allele predicted to be nonfunctional. All theoretical DNA codons are shown regardless of if they appear in any tRNAs in any species.

Codon (DNA sequence in tRNA gene)	Amino acid	Number of tRNA genes		
		<i>C. elegans</i>	<i>C. briggsae</i>	<i>C. tropicalis</i>
AGC	Ala	23	28	20
CGC	Ala	4	7	5
GGC	Ala	0	0	0
TGC	Ala	9	12	11
ACG	Arg	19	23	18
CCG	Arg	2	1	1
CCT	Arg	4	2	3
GCG	Arg	0	0	0
TCG	Arg	10	10	7
TCT	Arg	8	11	9
ATT	Asn	0	0	0
GTT	Asn	20	22	16
ATC	Asp	0	1	0
GTC	Asp	28	35	24
ACA	Cys	1	0	0
GCA	Cys	13	16	10
CTA	SeC	0	0	0
TCA	SeC	1	2	1
TTA	SeC	0	0	0
CTG	Gln	7	11	7
TTG	Gln	21	25	20
CTC	Glu	24	35	24
TTC	Glu	17	19	15
ACC	Gly	0	0	0
CCC	Gly	3	6	4
GCC	Gly	15	20	13
TCC	Gly	36	60	37
ATG	His	0	0	0
GTG	His	19	18	14
AAT	Ile	21	24	20

GAT	Ile	0	0	0
TAT	Ile	7	25	15
AAG	Leu	20	25	18
CAA	Leu	7	12	7
CAG	Leu	6	7	5
GAG	Leu	0	0	0
TAA	Leu	3	6	5
TAG	Leu	3	7	4
CTT	Lys	34	37	25
TTT	Lys	15	17	12
CAT	Met	10	13	11
CAT	iMet	9	10	9
AAA	Phe	0	0	1
GAA	Phe	15	21	14
AGG	Pro	7	6	8
CGG	Pro	4	4	5
GGG	Pro	0	1	1
TGG	Pro	32	47	38
ACT	Ser	0	0	1
AGA	Ser	15	16	12
CGA	Ser	6	8	5
GCT	Ser	10	9	6
GGA	Ser	0	0	0
TGA	Ser	9	7	7
AGT	Thr	17	20	16
CGT	Thr	7	7	5
GGT	Thr	0	0	0
TGT	Thr	12	11	8
CCA	Trp	12	14	9
ATA	Tyr	0	0	0
GTA	Tyr	19	21	12
AAC	Val	20	24	18
CAC	Val	6	6	7
GAC	Val	0	0	0
TAC	Val	5	5	7

Table S3. Metrics relating to the allele frequency spectra of genetic variants in all protein-coding exons, putatively functional tRNA genes, and putatively nonfunctional tRNA genes (the latter two groups comprise all sequences identified by tRNAScan-SE; see Methods for their definitions). All statistical test p-values shown were corrected for performing two pairwise tests in each case. Abbreviations used in table: MAF, minor allele frequency; SNV, single nucleotide variant; INDEL, insertion-deletion polymorphism.

		<i>C. elegans</i>			<i>C. briggsae</i>			<i>C. tropicalis</i>		
		Protein coding exons	tRNAs	Non-functional tRNAs	Protein coding exons	tRNAs	Non-functional tRNAs	Protein coding exons	tRNAs	Non-functional tRNAs
Singleton proportion related	Number variants (SNVs and INDELs)	859587	1360	385	829946	3262	618	626686	794	146
	Number singletons (% singletons)	246591 (28.7%)	683 (50.2%)	106 (27.5%)	307917 (37.1%)	1738 (53.3%)	213 (34.5%)	148307 (23.7%)	388 (48.9%)	48 (32.9%)
	χ^2 test results, protein-coding exons vs. tRNAs (functional only)	$\chi^2 = 855516, df = 1$ $p = 1 \times 10^{-92889}$			$\chi^2 = 820211, df = 1,$ $p = 1 \times 10^{-89056}$			$\chi^2 = 624308, df = 1$ $p = 1 \times 10^{-67786}$		
	χ^2 test results, protein-coding exons vs. functional and nonfunctional tRNAs	$\chi^2 = 854366, df = 1$ $p = 1 \times 10^{-92764}$			$\chi^2 = 818378, df = 1$ $p = 1 \times 10^{-88857}$			$\chi^2 = 623871, df = 1$ $p = 1 \times 10^{-67739}$		
Distribution central tendency related	Median MAF	0.0073	0.0015	0.0073	0.0042	0.0014	0.0042	0.0225	0.0032	0.0064
	Mann-Whitney test results, protein-coding exons vs. tRNAs (functional only)	$W = 404758843.5$ $p = 3 \times 10^{-88}$			$W = 1048948387.5$ $p = 9 \times 10^{-116}$			$W = 141744761$ $p = 2 \times 10^{-99}$		
	Mann-Whitney test results, protein-coding exons vs. functional and nonfunctional tRNAs	$W = 563015113$ $p = 1 \times 10^{-74}$			$W = 1308498571.5$ $p = 1 \times 10^{-95}$			$W = 17985348$ $p = 2 \times 10^{-96}$		

Table S4. Tajima's D in protein-coding exons, putatively functional tRNAs, and putatively nonfunctional tRNAs (the latter two groups comprise all sequences identified by tRNAScan-SE; see Methods for their definitions). Tajima's *D* was calculated with several input sets of variants, as noted in the table. Abbreviations used in table: INDEL, insertion-deletion polymorphism

Variant inclusion criteria	Metric	<i>C. elegans</i>			<i>C. briggsae</i>			<i>C. tropicalis</i>		
		Protein coding exons	tRNAs	Non-functional tRNAs	Protein coding exons	tRNAs	Non-functional tRNAs	Protein coding exons	tRNAs	Non-functional tRNAs
>=90% strains have genotype, biallelic sites including INDELS	N variants included	766486	1327	346	753207	3106	542	583407	784	130
	Tajima's <i>D</i>	-1.10	-2.15	-1.41	-1.49	-2.32	-1.56	-0.41	-2.24	-1.03
>=90% strains have genotype, biallelic sites excluding INDELS	N variants included	707446	1096	294	693696	2587	480	545587	698	106
	Tajima's <i>D</i>	-1.07	-2.12	-1.35	-1.45	-2.30	-1.48	-0.37	-2.21	-1.02
>=80% strains have genotype, biallelic sites including INDELS	N variants included	807038	1347	367	787664	3182	590	605533	788	133
	Tajima's <i>D</i>	-1.08	-2.14	-1.44	-1.48	-2.32	-1.52	-0.38	-2.24	-1.03
>=80% strains have genotype, biallelic sites excluding INDELS	N variants included	743117	1112	308	724103	2651	521	565449	699	108
	Tajima's <i>D</i>	-1.04	-2.11	-1.37	-1.44	-2.30	-1.43	-0.35	-2.21	-1.03

Table S5. All examples of putative isotype switching observed in this study where isotype switching is fixed in the species at this gene. Each gene with an isotype switching allele with RNA folding (Infernal) score > 25 is shown. Full per-allele information is available with the source data provided with this study.

Species	tRNA gene name	tRNA backbone (predicted by tRNAscan-SE)	Amino acid based on codon	Codon	N alleles at this gene with isotype switch	Total number alleles at this gene	RNA folding score of isotype switch alleles (Infernal, median)	N strains with isotype switch allele(s)	Switch allele freq. (of strains with non-missing genotypes here)	Switch allele freq. (of all strains)
<i>C. elegans</i>	II.trna16	Ala	Thr	AGT	1	1	74.8	684	100%	100%
	IV.trna13	Arg	Lys	CTT	3	3	26.1	194	94.63%	28.36%
	IV.trna74	Arg	Lys	CTT	1	1	26.1	684	100%	100%
	X.trna12	Arg	Thr	TGT	3	3	37.7	658	100%	96.20%
	X.trna292	His	Thr	TGT	1	1	48.1	684	100%	100%
	X.trna259	Ile	Lys	CTT	2	2	27.6	8	1.17%	1.17%
	I.trna10	Leu	Ile	TAT	1	1	84.7	684	100%	100%
	I.trna18	Leu	Ile	TAT	3	3	76.9	675	100%	98.68%
	I.trna23	Leu	Gly	CCC	1	1	79.1	684	100%	100%
	I.trna54	Leu	Gly	CCC	2	2	69.25	635	100%	92.84%
	I.trna55	Leu	Gly	CCC	2	2	79.15	681	100%	99.56%
	I.trna57	Leu	Ile	TAT	1	1	84.7	684	100%	100%
	III.trna60	Leu	Arg	CCT	1	1	68.7	684	100%	100%
	III.trna61	Leu	Lys	CTT	3	3	70.2	679	100%	99.27%
	IV.trna25	Leu	His	GTG	2	2	78.7	684	100%	100%
X.trna227	Leu	Ile	TAT	4	4	77.85	682	100%	99.71%	
<i>C. briggsae</i>	I.trna27	Ala	Leu	TAG	5	5	49.3	502	100%	69.82%
	I.trna116	Leu	Ile	TAT	6	6	76.9	704	100%	97.91%
	I.trna119	Leu	Ile	TAT	6	6	71.5	712	100%	99.03%
	I.trna120	Leu	Ile	TAT	6	6	78.1	696	100%	96.80%
	I.trna121	Leu	Ile	TAT	2	2	75.55	710	100%	98.75%

	I.trna122	Leu	Ile	TAT	4	4	76.05	714	100%	99.30%
	I.trna123	Leu	Ile	TAT	4	4	74.65	667	100%	92.77%
	I.trna124	Leu	Ile	TAT	1	1	79.4	719	100%	100%
	I.trna28	Leu	Ile	TAT	2	2	65.7	719	100%	100%
	I.trna29	Leu	Ile	TAT	4	4	70.65	669	100%	93.05%
	I.trna30	Leu	Ile	TAT	4	4	76.85	711	100%	98.89%
	I.trna31	Leu	Ile	TAT	3	3	69.5	687	100%	95.55%
	I.trna32	Leu	Ile	TAT	1	1	59.4	719	100%	100%
	I.trna33	Leu	Ile	TAT	2	2	64.35	714	100%	99.30%
	I.trna34	Leu	Ile	TAT	1	1	80.2	719	100%	100%
	I.trna35	Leu	Ile	TAT	8	8	73.65	675	100%	93.88%
	I.trna36	Leu	Ile	TAT	1	1	82.7	719	100%	100%
	I.trna37	Leu	Ile	TAT	3	3	80.2	710	100%	98.75%
	I.trna38	Leu	Ile	TAT	4	4	66.7	719	100%	100%
	I.trna39	Leu	Ile	TAT	3	3	65.2	715	100%	99.44%
	II.trna104	Leu	Gly	CCC	4	4	78.15	599	100%	83.31%
	II.trna17	Leu	Gly	CCC	2	2	80.35	717	100%	99.72%
	II.trna18	Leu	Gly	CCC	3	3	74	714	100%	99.30%
	II.trna19	Leu	Gly	CCC	7	7	76.6	688	100%	95.69%
	III.trna21	Ser	Leu	CAA	7	7	55.5	709	100%	98.61%
<i>C. tropicalis</i>	I.trna15	Leu	Ile	TAT	2	2	76.4	620	100%	99.68%
	I.trna16	Leu	Ile	TAT	1	1	78.5	622	100%	100%
	I.trna17	Leu	Ile	TAT	3	3	77.4	619	100%	99.52%
	I.trna20	Leu	Ile	TAT	1	1	78.5	622	100%	100%
	I.trna21	Leu	Ile	TAT	2	2	80.7	622	100%	100%
	I.trna68	Leu	Ile	TAT	1	1	75.8	622	100%	100%
	I.trna70	Leu	Ile	TAT	1	1	78.5	622	100%	100%
	I.trna71	Leu	Ile	TAT	3	3	76	620	100%	99.68%
	I.trna72	Leu	Ile	TAT	1	1	78.5	622	100%	100%
	I.trna73	Leu	Ile	TAT	2	2	74.3	620	100%	99.68%
	I.trna74	Leu	Ile	TAT	1	1	78.5	622	100%	100%
	I.trna77	Leu	Ile	TAT	2	2	76.45	619	100%	99.52%
	II.trna88	Leu	Gly	CCC	1	1	75.4	622	100%	100%
	II.trna89	Leu	Gly	CCC	1	1	69.9	622	100%	100%

II.trna90	Leu	Gly	CCC	2	2	66.7	616	100%	99.04%
II.trna91	Leu	Gly	CCC	1	1	69.9	622	100%	100%
I.trna22	Ser	Phe	AAA	1	1	27.9	622	100%	100%
I.trna4	Thr	Val	CAC	1	1	52	622	100%	100%
I.trna66	Thr	Pro	GGG	2	2	44.4	621	100%	99.84%
I.trna92	Thr	Leu	TAA	3	3	53.7	617	100%	99.20%
III.trna12	Trp	Arg	CCT	1	1	51.5	622	100%	100%
III.trna54	Trp	Gly	GCC	1	1	52.7	622	100%	100%
III.trna55	Trp	Arg	CCT	1	1	62	622	100%	100%
IV.trna13	Trp	Glu	CTC	1	1	63.5	622	100%	100%

Table S6. All examples of putative isotype switching observed in this study where isotype switching is polymorphic in the species at this gene. Each gene with an isotype switching allele with RNA folding (Infernal) score > 25 is shown. Full per-allele information is available with the source data provided with this study.

Species	tRNA gene name	tRNA backbone (predicted by tRNAScan-SE)	Amino acid based on codon	Codon	N alleles at this gene with isotype switch	Total number alleles at this gene	RNA folding score of isotype switch alleles (Infernal, median)	N strains with isotype switch allele(s)	Switch allele freq. (of strains with non-missing genotype s here)	Switch allele freq. (of all strains)
<i>C. elegans</i>	X.trna140	Glu	Lys	CTT	1	3	70.4	2	0.29%	0.29%
	X.trna23	Met	Thr	CGT	1	4	77	1	0.15%	0.15%
	V.trna12	Thr	Lys	TTT	2	5	72.35	2	0.29%	0.29%
	X.trna13	Thr	Lys, Arg	CTT, CCT	2	4	42	323	67.71%	47.22%
	X.trna19	Thr	Ile	AAT	1	4	78.8	5	0.75%	0.73%
	X.trna199	Thr	Ile	AAT	1	3	76.9	1	0.15%	0.15%
	X.trna299	Thr	Ile	AAT	1	3	85	1	0.15%	0.15%
	IV.trna33	Trp	Sup	TCA	1	4	71.5	1	0.15%	0.15%
	V.trna22	Val	Met	CAT	2	6	40	546	82.23%	79.82%
	III.trna7	iMet	Ile	AAT	1	2	61.2	1	0.15%	0.15%
<i>C. briggsae</i>	X.trna158	Arg	His	ATG	1	7	39.2	1	0.15%	0.14%
	III.trna33	Asn	Lys	TTT	1	6	73.5	1	0.14%	0.14%
	IV.trna54	Asn, Trp, Tyr	Sup, Trp	TCA, CCA	5	12	26.1	339	56.69%	47.15%
	V.trna30	Asn	His	GTG	1	5	46.8	2	0.33%	0.28%
	IV.trna1	Asp	Gly	GCC	1	14	46.8	1	0.17%	0.14%
	X.trna47	Gln	Sup	TTA	1	5	45	2	0.28%	0.28%
	II.trna110	Glu	Arg, Undet	TCT, NNN	1	8	50.55	18	2.80%	2.50%
	II.trna38	Glu	Sup	TCA	1	6	64.6	1	0.14%	0.14%
	I.trna134	Gly	Arg	TCT	1	4	71.7	1	0.14%	0.14%
	I.trna49	Gly	Glu	TTC	1	9	67.2	1	0.14%	0.14%
	II.trna81	Gly	Ser, Undet	CGA, NNN	1	4	47.55	10	1.41%	1.39%
	V.trna134	Gly	Ser	GCT	1	10	77.2	1	0.14%	0.14%
	X.trna117	Gly	Arg	TCT	1	7	45.9	1	0.15%	0.14%
	X.trna133	Ile	Val	TAC	1	8	75.4	1	0.14%	0.14%

	II.trna90	Lys	Sup	CTA	1	5	86	21	2.95%	2.92%
	V.trna15	Pro	Asn	ATT	1	2	56.3	1	0.15%	0.14%
	I.trna101	Thr	Ile	AAT	1	4	84.4	1	0.14%	0.14%
	I.trna129	Thr	Ile	AAT	1	4	83	2	0.28%	0.28%
	I.trna47	Thr	Ile	AAT	1	4	85	2	0.28%	0.28%
	I.trna65	Thr	Lys	TTT	2	7	72.4	4	0.57%	0.56%
	II.trna44	Thr	Ser	AGA	1	3	85.3	2	0.28%	0.28%
	II.trna67	Thr	Ile	AAT	1	6	84.6	12	1.74%	1.67%
	II.trna86	Thr	Ile	AAT	1	2	83	1	0.14%	0.14%
	II.trna89	Thr	Ile	AAT	2	6	86.75	5	0.70%	0.70%
	III.trna113	Thr	Ile	AAT	1	4	79.1	5	0.72%	0.70%
	III.trna12	Thr	Ile	AAT	1	4	79.1	1	0.14%	0.14%
	III.trna9	Thr	Ile	AAT	1	3	84.4	1	0.14%	0.14%
	X.trna105	Thr	Ile	AAT	1	4	79.6	1	0.14%	0.14%
	X.trna123	Thr	Tyr	GTA	1	6	60.9	1	0.14%	0.14%
	IV.trna88	Trp	Sup	TCA	1	4	40.3	4	0.63%	0.56%
	IV.trna89	Trp	Gly	TCC	1	4	42.6	1	0.15%	0.14%
	I.trna54	iMet	Leu	CAA	1	9	61.2	1	0.14%	0.14%
	III.trna106	iMet	Trp	CCA	1	5	43	18	2.53%	2.50%
<i>C. tropicalis</i>	II.trna100	Gly	Arg	TCG	1	2	71.7	2	0.33%	0.32%
	II.trna111	Gly	Ser	GCT	1	6	77.2	1	0.16%	0.16%
	I.trna54	Lys	Sup	TTA	1	3	72.5	3	0.48%	0.48%
	V.trna80	Met	Ile	TAT	1	4	77	1	0.16%	0.16%
	II.trna36	Thr	Lys	TTT	1	4	59.2	2	0.32%	0.32%
	II.trna76	Thr	Ile	AAT	1	2	85	1	0.16%	0.16%
	I.trna19	Trp	Arg	ACG	1	3	39.1	1	0.16%	0.16%

Table S7. Comparison of distribution of protein-coding and tRNA genes within hyperdivergent regions in *C. elegans*. The same set of protein-coding genes was used for each tRNA gene set comparison and the adjusted p-values correct for these three tests. 'All', putatively functional and putatively nonfunctional/pseudogenized tRNA genes; 'Func.', putatively functional tRNA genes; 'Pseud.', putatively non-functional tRNA genes (the latter two comprise all sequences in the 'All' category as identified by tRNAScan-SE; see Methods for their definitions). N, number; Prop., proportion; 'hyperdivergent' means overlapping a region hyperdivergent in any wild strain in the population per CaeNDR.

tRNA					Protein-coding genes				χ^2 test results			
tRNA gene set	N total	N hyper-divergent	N non-hyper-divergent	Prop. hyper-divergent	N total	N hyper-divergent	N non-hyper-divergent	Prop. hyper-divergent	χ^2	d.f.	P value (uncorrected)	P value (corrected)
All	721	124	597	0.17	19972	4749	15223	0.24	16.39	1	5.21E-05	0.0002
Func.	620	85	535	0.14					33.38	1	7.59E-09	2.28E-08
Pseud.	101	39	62	0.39					11.37	1	0.0007	0.0022