

Systematic identification of tRNA genes in *Drosophila melanogaster*

Steven J Marygold^{1§}, Patricia P Chan², Todd M Lowe²

¹FlyBase, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 3DY, UK

²Department of Biomolecular Engineering, University of California Santa Cruz, CA 95064, USA

[§]To whom correspondence should be addressed: sjm41@cam.ac.uk

Abstract

Transfer RNAs (tRNAs) are ubiquitous adapter molecules that link specific codons in messenger RNA (mRNA) with their corresponding amino acids during protein synthesis. The tRNA genes of *Drosophila* have been investigated for over half a century but have lacked systematic identification and nomenclature. Here, we review and integrate data within FlyBase and the Genomic tRNA Database (GtRNAdb) to identify the full complement of tRNA genes in the *D. melanogaster* nuclear and mitochondrial genomes. We apply a logical and informative nomenclature to all tRNA genes, and provide an overview of their characteristics and genomic features.

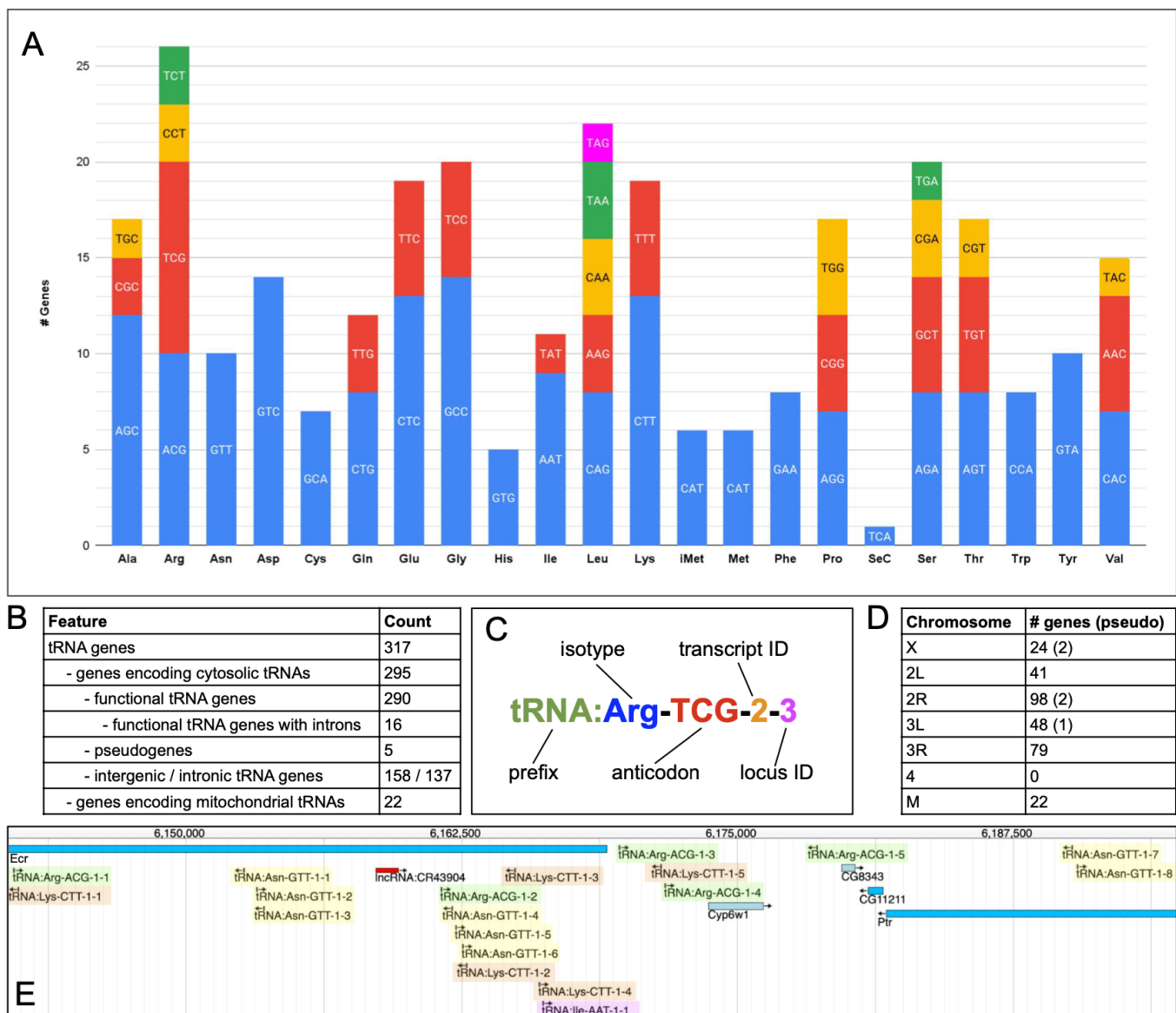


Figure 1. *Drosophila* tRNA genes

(A) Isoacceptor counts of functional cytosolic tRNA genes in *Drosophila* - pseudogenes and mitochondrial tRNA genes are excluded. Colors distinguish distinct anticodons within each isoacceptor family. Counts for initiator (iMet) and elongator tRNA:Met are shown separately. (B) Summary statistics for all *Drosophila* tRNA genes. (C) Example of tRNA gene nomenclature syntax. (D) Distribution of tRNA genes across major chromosomal scaffolds; pseudogene numbers are in parentheses. 'M' = mitochondrial genome. (E) Example of a tRNA gene cluster from the 2R:6,144,000..6,195,000 genomic region, corresponding to cytological region 42A shown in Figure 3 of Kubli 1982. (Notably, the only significant change to tRNA annotations in this region is the addition of tRNA:Arg-ACG-1-4 in the current version.) Green highlight: tRNA:Arg-ACG-1 isodecoders; orange highlight: tRNA:Lys-CTT-1; yellow highlight: tRNA:Asn-GTT-1; magenta highlight: tRNA:Ile-AAT-1; blue rectangles: protein-coding genes; red rectangle is a lncRNA gene.

Description

tRNAs are universal to all cellular life and provide the essential molecular link between mRNA codons and their corresponding amino acids during translation (reviewed by Suzuki 2021). The main functional regions in a tRNA are the anticodon triplet, which base pairs with mRNA codons, and the 3' end to which the cognate amino acid is attached. Codon degeneracy for the 21 amino acids (20 standard amino acids plus selenocysteine) means that up to six tRNAs with distinct anticodons ('isoacceptors') are required depending on the amino acid. tRNA diversity is further increased through the existence of tRNAs that share the same anticodon but differ in the sequence of their body structure (Goodenbour and Pan 2006). Such 'isodecoders' may differ from each other by just one or several nucleotides. Moreover, each specific isodecoder sequence can be present in multiple copies within a genome. This combination of diversity and redundancy results in eukaryotic nuclear genomes having hundreds of genes encoding tRNAs functioning in cytosolic translation (cytosolic tRNAs). An additional set of tRNAs functioning in mitochondria are encoded by the mitochondrial genome of eukaryotes: in vertebrates and many other metazoa, there are often 22 mitochondrial tRNA genes with tRNA:Leu and tRNA:Ser represented by two different isoacceptors.

The tRNAs and tRNA genes of *Drosophila melanogaster* (hereafter 'Drosophila') have been investigated for over half a century. Early work identified the sequences and locations of many individual cytosolic tRNA genes, demonstrating that several isoacceptor families appear as clusters in cytological (polytene chromosome) views of the genome (reviewed in Kubli 1982 and Sharp *et al.* 1983). Subsets of mitochondrial tRNA genes were also reported (de Bruijn 1983; Garesse 1988). Subsequently, the analysis of the sequenced *Drosophila* genome predicted 292 cytosolic tRNA genes (Adams *et al.* 2000). This number was later refined in release 4 of the genome to 297 cytosolic tRNA genes, including four pseudogenes (*Drosophila* 12 Genomes Consortium 2007; Bergman and Ardell 2014). However, none of these studies fully classified or specifically named tRNA genes by their molecular features, and the genomic data were not fully rationalized with the earlier information on tRNAs present within FlyBase, the primary database for *Drosophila* research (Larkin *et al.* 2021).

We revisited cytosolic tRNA gene annotations in the current version of the *Drosophila* genome (release 6), integrating gene predictions from the Genomic tRNA database (GtRNAdb; Chan and Lowe 2016; Chan *et al.* 2021) with existing data within FlyBase. We find there are a total of 295 cytosolic tRNA genes, of which 289 encode tRNAs decoding the 20 standard amino acids, one encodes a selenocysteine tRNA, and five are classified as tRNA-like genes/pseudogenes (Figure 1A, 1B; Extended Data Table 1). Importantly, this rationalization exercise resulted in improvements within both databases. For example, ~20 records pertaining to unlocalized tRNA genes were merged/deleted within FlyBase, two tRNAs with undetermined isotypes within the GtRNAdb were corrected, and pseudogene classifications within both databases were made consistent.

Prior to our analysis, greater than 50% of *Drosophila* cytosolic tRNA genes were unnamed in FlyBase, while the named genes used an ambiguous and esoteric nomenclature incorporating the single letter amino acid code and cytogenetic map information but lacking anticodon information. We therefore implemented the logical and systematic nomenclature used by the GtRNAdb within FlyBase (Figure 1C; Extended Data Table 1). This syntax comprises the 3-letter code of the cognate amino acid (isotype), the anticodon triplet, a number identifying each unique transcript (isodecoder) sequence, followed by a second number to specify each copy (locus) of that sequence within the genome. This is preceded by the standard 'tRNA:' prefix used for tRNA genes in FlyBase. For example, *tRNA:Arg-TCG-2-3* is the systematic name given to the gene encoding the third copy of the second unique transcript of tRNA:Arg-TCG. tRNA pseudogenes are named using the standard FlyBase syntax for pseudogenes, with a greek Psi character appended to the gene name, e.g. *tRNA:His-GTG-2-1Ψ*.

Mitochondrial tRNA genes are not currently included within the GtRNAdb. We therefore compared existing FlyBase annotations against *Drosophila* entries in the mitotRNAdb (Jühling *et al.* 2009) and verified these using the tRNAscan-SE program (Chan *et al.* 2021). This resulted in minor edits to five mitochondrial tRNA sequences in FlyBase. We also applied the standardized nomenclature to the 22 mitochondrial tRNA genes, including the standard 'mt:' prefix used for genes of the mitochondrial genome in FlyBase (Figure 1B; Extended Data Table 2).

Extended Data Tables 1 and 2 provide detailed information on all the cytosolic and mitochondrial tRNA genes, respectively. Among the functional, cytosolic tRNAs decoding the 20 standard amino acids, each isoacceptor is encoded by between five (tRNA:His) and 26 (tRNA:Arg) genes, with the number of distinct anticodons in each isoacceptor family ranging from one (e.g. tRNA:Asn-GTT) to five (tRNA:Leu-CAG, tRNA:Leu-AAG, tRNA:Leu-CAA, tRNA:Leu-TAA, tRNA:Leu-TAG). Up to four distinct transcript sequences exist for a given anticodon (as is the case for tRNA:Arg-TCG, tRNA:Cys-GCA, tRNA:Gln-CTG and tRNA:Leu-TAA), and a given transcript sequence may be present in up to 13 exact gene copies (tRNA:Gly-GCC-1 and tRNA:Lys-CTT-1). Overall, there are 44 different isoacceptors and 84 distinct tRNA transcripts encoded by the *Drosophila* nuclear genome. Cytosolic tRNA genes are present on all major chromosome arms (Figure 1D) and are frequently found within clusters, harboring members of either the same or different isoacceptor families (Figure 1E; Kubli 1982; Phillips and Ardell 2021). 46% of cytosolic tRNA genes are located within introns of protein-coding genes, with the remainder being intergenic. A minority (5%) of cytosolic tRNA genes contain an intron, namely two tRNA:Ile-TAT, four tRNA:Leu-CAA and ten tRNA:Tyr-GTA genes (Figure 1B; Bergman and Ardell 2014). These characteristics are largely comparable with the cytosolic tRNA gene complement of other metazoa (<http://gtRNadb.ucsc.edu/>). However, the *Drosophila* genome is distinguished by its relative paucity of non-functional tRNA-like genes: <2% of *Drosophila* tRNA genes are classed as pseudogenes or repetitive element derivatives, compared to 20% in *C. elegans*, 30% in humans or >99% in rodents.

Notably, this project enabled several additional improvements to the representation of tRNAs within FlyBase. All functional (Gene Ontology) annotations were reviewed and revised as necessary. Reciprocal links between tRNA gene/transcript reports in FlyBase and corresponding pages at the GtRNadb and RNACentral (RNACentral Consortium 2021) have been established, and 2D structural images for *Drosophila* tRNAs (Sweeney *et al.* 2021) have been added. Finally, a ‘Gene Group’ report (Attrill *et al.* 2016) for the *Drosophila* tRNA genes has been generated (<https://flybase.org/reports/FBgg0000459.html>), which provides easy access to all classes of tRNA gene and their associated data within FlyBase.

In conclusion, we have generated definitive sets of cytosolic and mitochondrial tRNA genes present in the *Drosophila* genome and implemented a systematic and informative nomenclature for them. The improved datasets are available from several databases, including FlyBase, GtRNadb, RNACentral, and the Alliance of Genome Resources (Alliance of Genome Resources Consortium 2022). Our work will facilitate further exploration of tRNA biology within *Drosophila* as well as new comparative studies with other species.

Methods

Data on cytosolic tRNA genes were accessed and downloaded from FlyBase (<http://flybase.org>) and the GtRNadb (<http://gtRNadb.ucsc.edu/>). Data were initially compared and rationalized between FlyBase release FB2015_04 and GtRNadb release 16. Necessary revisions were made in subsequent database releases, and the data presented herein are from FlyBase release FB2022_02 and GtRNadb release 19 (which uses tRNAscan-SE 2.0). tRNAscan-SE uses multiple score thresholds to discriminate functional tRNA genes from likely non-functional tRNA-like genes/pseudogenes based on their primary sequence, secondary structure and isotype-specific models (Chan *et al.* 2021).

Data on mitochondrial tRNA genes were accessed and downloaded from FlyBase (<http://flybase.org>) and mitotRNadb (<http://mttrna.bioinf.uni-leipzig.de/mtDataOutput/>). Additionally, the *Drosophila* mitochondrial genome (RefSeq NC_024511) was used as a query sequence in tRNAscan-SE 2.0 (<http://trna.ucsc.edu/tRNAscan-SE/>) with parameters ‘other mitochondrial’ as the sequence source, ‘Invertebrate Mito’ as the genetic code and a score cutoff of zero. Necessary revisions were made in FlyBase and data presented herein are from FlyBase release FB2022_02.

Reagents

Extended Data

Description: Cytosolic tRNA genes. Resource Type: Dataset. File: [Extended Data Table 1.xlsx](#). DOI: [10.22002/D1.20161](https://doi.org/10.22002/D1.20161)

Description: Mitochondrial tRNA genes. Resource Type: Dataset. File: [Extended Data Table 2.xlsx](#). DOI: [10.22002/D1.20162](https://doi.org/10.22002/D1.20162)

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287: 2185-95. PubMed ID: [10731132](https://pubmed.ncbi.nlm.nih.gov/10731132/)

Alliance of Genome Resources Consortium. 2022. Harmonizing model organism data in the Alliance of Genome Resources. *Genetics* 220: iyac022. PubMed ID: [35380658](https://pubmed.ncbi.nlm.nih.gov/35380658/)

- Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ, FlyBase Consortium. 2016. FlyBase: establishing a Gene Group resource for *Drosophila melanogaster*. *Nucleic Acids Res* 44: D786-92. PubMed ID: [26467478](#)
- Bergman C, Ardell D. 2014. Methods for nuclear tRNA gene predictions for 12 species in the genus *Drosophila*. figshare. Journal contribution. DOI: [10.6084/m9.figshare.1233438.v1](#)
- Chan PP, Lin BY, Mak AJ, Lowe TM. 2021. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res* 49: 9077-9096. PubMed ID: [34417604](#)
- Chan PP, Lowe TM. 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* 44: D184-9. PubMed ID: [26673694](#)
- de Bruijn MH. 1983. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. *Nature* 304: 234-41. PubMed ID: [6408489](#)
- Drosophila* 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203-18. PubMed ID: [17994087](#)
- Garesse R. 1988. *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary considerations. *Genetics* 118: 649-63. PubMed ID: [3130291](#)
- Goodenbour JM, Pan T. 2006. Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res* 34: 6137-46. PubMed ID: [17088292](#)
- Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J. 2009. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* 37: D159-62. PubMed ID: [18957446](#)
- Kubli E. 1982. The genetics of transfer RNA in *Drosophila*. *Adv Genet* 21: 123-72. PubMed ID: [6800234](#)
- Larkin A, Marygold SJ, Antonazzo G, Attrill H, Dos Santos G, Garapati PV, Goodman JL, Gramates LS, Millburn G, Strelets VB, Tabone CJ, Thurmond J, FlyBase Consortium. 2021. FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res* 49: D899-D907. PubMed ID: [33219682](#)
- Phillips JB, Ardell DH. 2021. Structural and Genetic Determinants of Convergence in the *Drosophila* tRNA Structure-Function Map. *J Mol Evol* 89: 103-116. PubMed ID: [33528599](#)
- RNAcentral Consortium. 2021. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res* 49: D212-D220. PubMed ID: [33106848](#)
- Sharp S, Cooley L, DeFranco D, Dingermann T, Söll D. 1983. Organization and expression of tRNA genes in *Drosophila melanogaster*. *Recent Results Cancer Res* 84: 1-14. PubMed ID: [6405456](#)
- Suzuki T. 2021. The expanding world of tRNA modifications and their disease relevance. *Nat Rev Mol Cell Biol* 22: 375-392. PubMed ID: [33658722](#)
- Sweeney BA, Hoksza D, Nawrocki EP, Ribas CE, Madeira F, Cannone JJ, Gutell R, Maddala A, Meade CD, Williams LD, Petrov AS, Chan PP, Lowe TM, Finn RD, Petrov AI. 2021. R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nat Commun* 12: 3494. PubMed ID: [34108470](#)

Funding: SJM is funded by a grant from the National Human Genome Research Institute, National Institutes of Health (U41HG000739) to Norbert Perrimon (PI), Nicholas Brown (co-PI). PPC and TML are funded by a grant from the National Human Genome Research Institute, National Institutes of Health (R01HG006753) to TML.

Author Contributions: Steven J Marygold: conceptualization, data curation, writing - original draft, visualization, methodology. Patricia P Chan: writing - review editing, validation, software, methodology. Todd M Lowe: funding acquisition, software, methodology, validation, writing - review editing.

Reviewed By: Anonymous

History: Received March 16, 2022 **Revision Received** April 19, 2022 **Accepted** April 19, 2022 **Published** May 1, 2022

Copyright: © 2022 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Marygold, SJ; Chan, PP; Lowe, TM (2022). Systematic identification of tRNA genes in *Drosophila melanogaster*. *microPublication Biology*. [10.17912/micropub.biology.000560](#)

