

Gene model for the ortholog of *dib* in *Drosophila cardini*

Parker Brooks¹, Pablo Chialvo¹, Clare Scott Chialvo^{1§}

¹Biology, Appalachian State University, Boone, North Carolina USA

[§]To whom correspondence should be addressed: chialvoch@appstate.edu

Abstract

We developed a gene model for the *disembodied* ortholog (*dib*) in the ASM1890373v1 Genome Assembly (GenBank Accession: [GCA_018903735.1](https://www.ncbi.nlm.nih.gov/nuccore/GCA_018903735.1)) of *Drosophila cardini*. This ortholog was characterized as part of a developing dataset for a comparative study of detoxification gene family evolution in the *immigrans-tripunctata* radiation of the genus *Drosophila* using an adapted Genomics Education Partnership gene annotation protocol for Course-based Undergraduate Research Experiences.

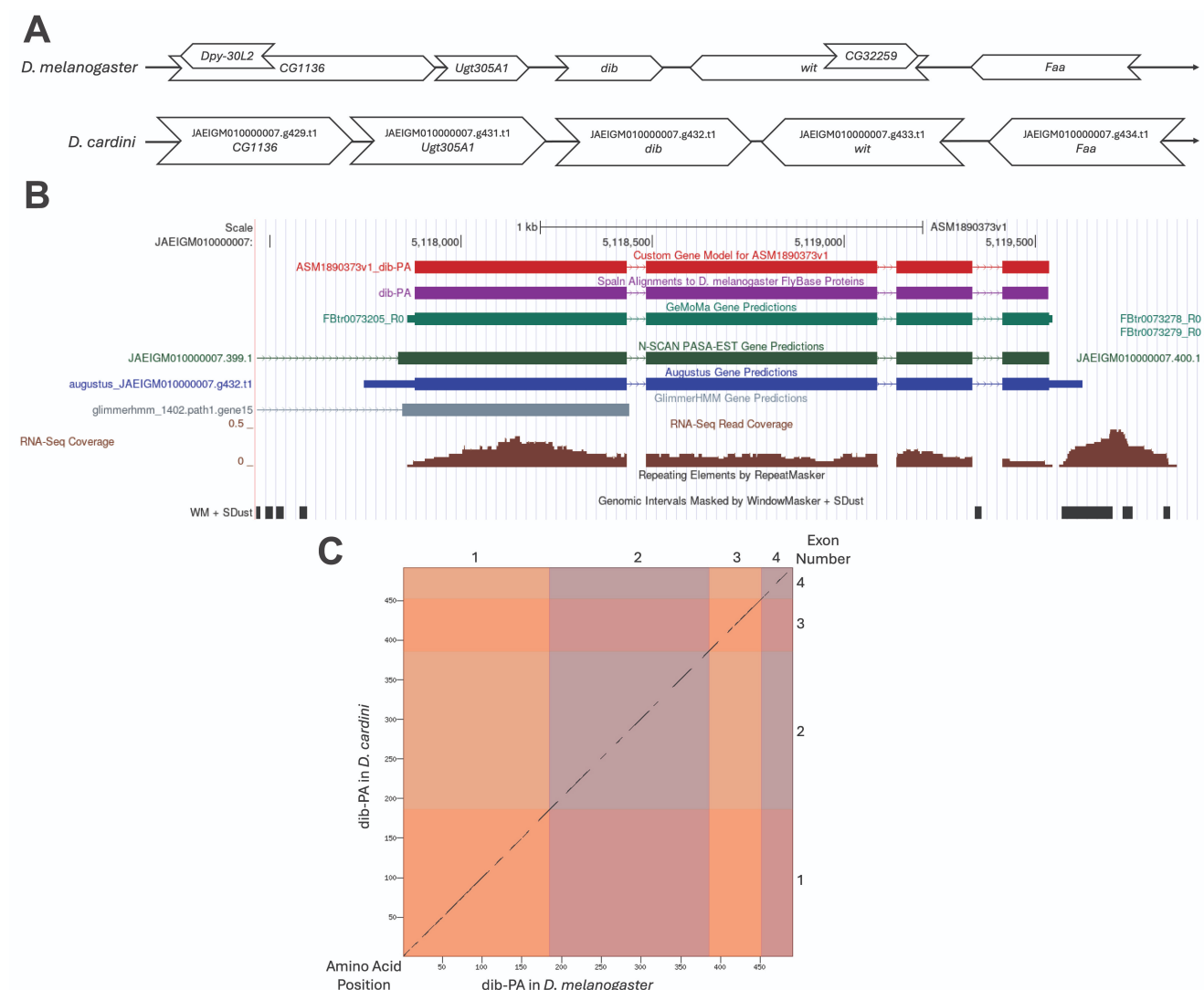


Figure 1. Genomic neighborhood and gene model for *dib* in *D. cardini*:

(A) Synteny comparison of the genomic neighborhoods for *dib* in *Drosophila melanogaster* and *D. cardini*. Thin underlying arrows indicate which DNA strand the target gene, *dib*, is located on in *D. melanogaster* (top) and *D. cardini* (bottom). The thin arrows pointing to the right indicate that *dib* is on the positive strand in both *D. melanogaster* and *D. cardini*. The wide gene arrows pointing in the same direction as *dib* are on the same strand relative to the thin underlying arrows, while wide gene arrows pointing in the opposite direction of *dib* are on the opposite strand relative to the thin underlying arrows. White gene arrows in *D. cardini* indicate orthology to the corresponding gene in *D. melanogaster*. Gene symbols given in the *D. cardini* gene arrows indicate the orthologous gene in *D. melanogaster*, while the locus identifiers are specific to *D. cardini*. (B) Gene Model in GEP UCSC Track Data Hub (Raney et al., 2014). The coding-regions of *dib* in *D. cardini* are displayed in the User Supplied Track (red); coding sequences (CDS) are depicted by thick

rectangles and introns by thin lines with arrows indicating the direction of transcription. Subsequent evidence tracks include Spaln of *D. melanogaster* Proteins (purple, alignment of Ref-Seq proteins from *D. melanogaster*), Coding Regions Predicted by Augustus (dark blue), GeMoMa (teal), and NSCAN PASA-EST (dark green), and RNA-Seq from mixed sex adult flies (brown; alignment of Illumina RNA-Seq reads from *D. cardini* – Erlenbach et al. 2023). (C) **Dot Plot of dib-PA in *D. melanogaster* (x-axis) vs. the orthologous peptide in *D. cardini* (y-axis).** Amino acid number is indicated along the left and bottom; CDS number is indicated along the top and right, and CDSs are also highlighted with alternating colors. Line breaks in the dot plot indicate areas of amino acid mismatch between species. We do not note any significant breaks.

Description

Introduction

This article reports a predicted gene model generated by undergraduate work using a structured gene model annotation protocol defined by the Genomics Education Partnership (GEP; thegep.org) for Course-based Undergraduate Research Experience (CURE). The following information in quotes may be repeated in other articles submitted by participants using the same GEP CURE protocol for annotating *Drosophila* species orthologs of *Drosophila melanogaster* detoxification genes.

“Within insects, detoxifying xenobiotics and host secondary metabolites is a three-phase process that involves functionalization, conjugation, and excretion of these compounds. Expansions of known detoxification gene families (*e.g.*, cytochrome P450s) are associated with diet breadth and insecticide resistance (Ranson et al., 2002; Després et al., 2007; Rane et al., 2016). With the increasing availability of high-quality genomes for non-model organisms, including *Drosophila* species beyond *D. melanogaster*, it is now possible to perform large scale comparative studies (Robinson et al., 2011; Kim et al., 2021; Threfall and Baxter 2021). Careful manual annotation and curation of gene models can improve upon computational gene predictions in non-model species, which aids the accuracy of studies on gene and genome evolution (Mudge and Harrow 2016; Tello-Ruiz et al., 2019). To aid in these annotations, the Genomics Education Partnership (thegep.org) developed a curriculum involving web-based tools that allow undergraduates to engage in authentic course-based research focused on manually annotating genes in non-model species (Rele et al., 2023). The orthologous gene models, including the one presented here, then provide a reliable basis for further evolutionary genomic analyses when made available to the scientific community. The gene ortholog described here in *D. cardini* for *disembodied* (*dib*), a member of the cytochrome P450 monooxygenase gene family, was characterized as part of a developing dataset for a comparative study of detoxification gene families in the *immigrans-tripunctata* radiation of the genus *Drosophila*.” (Williams et al, 2026)

“In the subgenus *Drosophila*, *D. cardini* Sturtevant 1916 is a member of the *cardini* subgroup in the *cardini* species group of the *immigrans-tripunctata* radiation (Heed & Krishnamurthy 1959; Bächli 2005). Species in the *cardini* subgroup are found in the mainland Neotropics, and the range of *D. cardini* extends from Florida to Brazil (Heed, 1962). Members of the *cardini* group primarily feed and develop on fruit and flowers (Markow & O’Grady 2008). However, *D. cardini* is also reported to feed on mushrooms and can tolerate the cyclopeptide toxin α -amanitin (Stump et al. 2011).” (Patel et al., 2026)

“Cytochrome P450 monooxygenases (CYPs) are a family of phase I detoxification enzymes that are found in almost all aerobic organisms and act by oxidizing compounds to make them more polar (Stegeman and Livingstone, 1998; Li et al., 2007). The enzymes in this family vary in both their substrate specificity and the range of metabolites that they produce (Rendic and Di Carlo 1997; Scott 1999). Furthermore, the substrate specificity of CYPs can be altered by a change in a single amino acid (Lindberg and Negishi, 1989).” (Patel et al., 2026)

disembodied (*dib*), a mitochondrial CYP whose expression is restricted to the prothoracic gland, is a member of the Halloween gene group (Gilbert 2004). These genes act in the biosynthetic pathway that produces 20-Hydroxyecdysone, a hormone that controls metamorphosis and development in arthropods (Gilbert and Warren 2005; Rewitz et al., 2006). Homozygous *dib* mutants die prior to completing embryogenesis and display abnormal cuticular development (Warren et al. 2002; Gilbert 2004).

We propose a gene model for the *D. cardini* ortholog of the *D. melanogaster* *disembodied* (*dib*) gene. The genomic region of the ortholog corresponds to the Augustus gene prediction JAEIGM010000007.g432.t1 in the ASM1890373v1 Genome Assembly of *D. cardini* ([GCA_018903735.1](https://doi.org/10.5061/dryad.hdr7sqvq2) – Kim et al., 2021). This model is based on the mixed sex adult RNA-Seq data from *D. cardini* (Erlenbach et al., 2023; <https://doi.org/10.5061/dryad.hdr7sqvq2>) and *dib* in *D. melanogaster* using FlyBase release FB2024_02 ([GCA_000001215.4](https://doi.org/10.5061/dryad.hdr7sqvq2); Gramates et al., 2022; Jenkins et al., 2022; Larkin et al., 2021).

Synteny

The reference gene, *dib*, occurs on chromosome 3L in *D. melanogaster* and is flanked upstream by UDP-glycosyltransferase family 305 member A1 (*Ugt305A1*), [CG1136](#), and *Dpy-30-like 2* ([Dpy-30L2](#)) which is nested in [CG1136](#) and downstream by *wishful thinking* (*wit*), [CG32259](#) (which is nested in *wit*), and *Fumarylacetoacetase* (*Faa*). The *tblastn* search of *D. melanogaster* *dib*-PA (query) against the *D. cardini* genome assembly (GenBank Accession:

[GCA_018903735.1](#); subject) placed the putative ortholog of *dib* within contig_2202 (JAEIGM010000007.1) which corresponds to Augustus gene prediction JAEIGM010000007.g432.t1 (E-value: 0.0; percent identity: 79.75% as determined by *blastp*). The putative ortholog is flanked upstream by the Augustus gene predictions JAEIGM010000007.g431.t1 and JAEIGM010000007.g429.t1, which correspond to *UGt305A1* and *CG1136* in *D. melanogaster* (E-value: 5e-160 and 0.0; identity: 45.3% and 65.92%, respectively, as determined by *blastp*; Figure 1A; Altschul et al., 1990). Augustus gene prediction JAEIGM010000007.g430.t1 did not map to gene expression data and was not included. The putative ortholog of *dib* is flanked downstream by the Augustus gene predictions JAEIGM010000007.g433.t1 and JAEIGM010000007.g434.t1, which correspond to *wit* and *Faa* in *D. melanogaster* (E-value: 0.0 and 0.0; identity: 74.81% and 87.77%, respectively, as determined by *blastp*). The putative ortholog assignment for *dib* in *D. cardini* is supported by the following evidence: The gene predictions surrounding the *dib* ortholog are orthologous to the genes at the same locus in *D. melanogaster*, gene expression data corresponds with each prediction, and local synteny is completely conserved, supported by E-values and percent identities, so we conclude that the Augustus prediction JAEIGM010000007.g432.t1 represents an ortholog of *dib* in *D. cardini* (Figure 1A).

Protein Model

dib in *D. cardini* has 4 coding sequences (CDS) within its genomic DNA sequence. The only unique protein sequence (dib-PA) is translated from 1 mRNA isoform (dib-RA; Figure 1B). Relative to the ortholog in *D. melanogaster*, the CDS number and protein isoform count are conserved. The sequence of dib-PA in *D. cardini* has 79.4% identity (91.0% similarity) with the protein-coding isoform dib-PA in *D. melanogaster*, as determined by *blastp* (Figure 1C). This level of divergence is not surprising given that *D. cardini* and *D. melanogaster* belong to two separate subgenera (*Drosophila* and *Sophophora* respectively) that diverged approximately 45-60 MYA (Russo et al., 1995; Tamura et al. 2004; Obbard et al., 2012). Coordinates of this curated gene model are archived in the CaltechDATA repository (see “Extended Data” section below).

Methods

The annotation methods used in this project are adapted from those described in Rele et al. (2023), which includes algorithms, database versions, and citations for the complete annotation process developed for the Pathways Project. The methods for the current project are detailed in brief below with notes on significant differences between this protocol and the one described in Rele et al. (2023). The students use the GEP instance of the UCSC Genome Browser v.435 (<https://gander.wustl.edu>; Kent WJ et al., 2002; Raney et al., 2024) to examine the genomic neighborhood of their reference detoxification gene in the *D. melanogaster* genome assembly (Aug. 2014; BDGP Release 6 + ISO1 MT/dm6). Students obtain the protein sequence for the *D. melanogaster* target gene for a given isoform and use a *tblastn* search of the sequence against their target *Drosophila* species genome assembly (*D. cardini* (GCA_018903735.1 - Kim et al., 2021)) on the NCBI BLAST server (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, Altschul et al., 1990) to identify the putative ortholog location. Students compare the genomic neighborhood of the putative ortholog to that of the reference gene in *D. melanogaster*. This local synteny analysis includes a minimum of two upstream and two downstream genes relative to the potential ortholog. As no RefSeq protein data is available for these species, comparisons are based on gene predictions that correlate with gene expression data in the putative ortholog neighborhood. Using the multiple alignment tracks feature in the Genome Browser, students examine other sets of genomic evidence, including Spaln alignment of *D. melanogaster* proteins, multiple gene prediction tracks (e.g., GeMoMa, Augustus, NSCAN PASA-EST), and RNA-Seq mixed seq adult expression data from the target species generated by Erlenbach et al. (2023; <https://doi.org/10.5061/dryad.hdr7sqvq2>). Information on the genomic structure information (e.g., CDSs, intron-exon number, number of isoforms) for the reference gene in *D. melanogaster* is retrieved using Gene Record Finder (<https://gander.wustl.edu/~wilson/dmelgenerecord/index.html>; Rele et al., 2023). To determine approximate splice sites within the target gene, a *tblastn* search using the CDSs from the *D. melanogaster* reference gene against the putative ortholog location (10kb up- and downstream of the target gene prediction). Coordinates of the CDS(s) are refined by examining aligned RNA-Seq data, identifying canonical splice site sequences, and ensuring the maintenance of an open reading frame. Students confirm the biological validity of their target gene model using the FlySeq Gene Model Checker (<https://gander2.wustl.edu/~wilson/genechecker-flyseq/>), which compares the hypothesized target gene model's structure and translated sequence against the *D. melanogaster* reference gene. At least two independent models for this gene are generated. These models are reconciled by a third independent researcher to produce the final model presented here. Note: comparison of 5' and 3' UTR sequence information is not included in this GEP CURE protocol.

Acknowledgements: We would like to thank Wilson Leung for developing and maintaining the technological infrastructure that was used to create this gene model and Laura K. Reed for overseeing the Genomics Education Partnership. Thank you to FlyBase for providing the definitive database for *Drosophila melanogaster* gene models.

Extended Data

Description: Zipped archive containing FASTA, PEP, and GFF files for *dib* model. Resource Type: Dataset. File: [Dcar_dib_model.tar.gz](#). DOI: [10.22002/pm412-0xx64](https://doi.org/10.22002/pm412-0xx64)

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215(3): 403-10. PubMed ID: [2231712](#)
- Bächli, G. (2005) Taxodros: The database on taxonomy of Drosophilidae, version February 2026, last accessed 28 May 2026. <https://taxodros.uzh.ch/>
- Després L, David JP, Gallet C. 2007. The evolutionary ecology of insect resistance to plant chemicals. *Trends Ecol Evol* 22(6): 298-307. PubMed ID: [17324485](#)
- Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al., MacCallum I. 2007. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* 450(7167): 203-18. PubMed ID: [17994087](#)
- Erlenbach T, Haynes L, Fish O, Beveridge J, Giambrone SA, Reed LK, Dyer KA, Scott Chialvo CH. 2023. Investigating the phylogenetic history of toxin tolerance in mushroom-feeding Drosophila. *Ecol Evol* 13(12): e10736. PubMed ID: [38099137](#)
- Gilbert LI. 2004. Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in Drosophila melanogaster. *Mol Cell Endocrinol* 215(1-2): 1-10. PubMed ID: [15026169](#)
- Gilbert LI, Warren JT. 2005. A molecular genetic approach to the biosynthesis of the insect steroid molting hormone. *Vitam Horm* 73: 31-57. PubMed ID: [16399407](#)
- Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, Dos Santos G, et al., the FlyBase Consortium. 2022. FlyBase: a guided tour of highlighted features. *Genetics* 220(4): 10.1093/genetics/iyac035. PubMed ID: [35266522](#)
- Heed, W.B., Krishnamurthy, N.B. (1959). Genetic studies on the cardini group of Drosophila in the West Indies. *University of Texas Publication* 5914, 155-179.
- Heed, W.B. (1962) Genetic characteristics of island populations. *University of Texas Publication* 6205, 173-206.
- Jenkins VK, Larkin A, Thurmond J, FlyBase Consortium. 2022. Using FlyBase: A Database of Drosophila Genes and Genetics. *Methods Mol Biol* 2540: 1-34. PubMed ID: [35980571](#)
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* 12(6): 996-1006. PubMed ID: [12045153](#)
- Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, et al., Petrov DA. 2021. Highly contiguous assemblies of 101 drosophilid genomes. *Elife* 10: 10.7554/eLife.66405. PubMed ID: [34279216](#)
- Larkin A, Marygold SJ, Antonazzo G, Attrill H, Dos Santos G, Garapati PV, et al., FlyBase Consortium. 2021. FlyBase: updates to the Drosophila melanogaster knowledge base. *Nucleic Acids Res* 49(D1): D899-D907. PubMed ID: [33219682](#)
- Li X, Schuler MA, Berenbaum MR. 2007. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annu Rev Entomol* 52: 231-53. PubMed ID: [16925478](#)
- Lindberg RL, Negishi M. 1989. Alteration of mouse cytochrome P450c11 substrate specificity by mutation of a single amino-acid residue. *Nature* 339(6226): 632-4. PubMed ID: [2733794](#)
- Markow TA, O'Grady P. 2008. Reproductive ecology of *Drosophila*. *Functional Ecology* 22: 747-759. DOI: [10.1111/j.1365-2435.2008.01457.x](#)
- Mudge JM, Harrow J. 2016. The state of play in higher eukaryote gene annotation. *Nat Rev Genet* 17(12): 758-772. PubMed ID: [27773922](#)
- Obbard DJ, MacLennan J, Kim KW, Rambaut A, O'Grady PM, Jiggins FM. 2012. Estimating divergence dates and substitution rates in the Drosophila phylogeny. *Mol Biol Evol* 29(11): 3459-73. PubMed ID: [22683811](#)
- Patel P, Chialvo P, Scott Chialvo C. 2026. Gene model for the ortholog of sad in Drosophila cardini. *MicroPubl Biol* 2026: 10.17912/micropub.biology.002229. PubMed ID: [42434616](#)
- Rane RV, Walsh TK, Pearce SL, Jermini LS, Gordon KH, Richards S, Oakeshott JG. 2016. Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? *Curr Opin Insect Sci* 13: 70-76. PubMed ID: [27436555](#)
- Raney BJ, Barber GP, Benet-Pagès A, Casper J, Clawson H, Cline MS, et al., Haeussler M. 2024. The UCSC Genome Browser database: 2024 update. *Nucleic Acids Res* 52(D1): D1082-D1088. PubMed ID: [37953330](#)
- Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, et al., Kent WJ. 2014. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 30(7): 1003-5. PubMed ID: [24227676](#)

- Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, et al., Feyereisen R. 2002. Evolution of supergene families associated with insecticide resistance. *Science* 298(5591): 179-81. PubMed ID: [12364796](#)
- Rele CP, Sandlin KM, Leung W, Reed LK. 2022. Manual annotation of *Drosophila* genes: a Genomics Education Partnership protocol. *F1000Res* 11: 1579. PubMed ID: [37854289](#)
- Rendic S, Di Carlo FJ. 1997. Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab Rev* 29(1-2): 413-580. PubMed ID: [9187528](#)
- Rewitz KF, Rybczynski R, Warren JT, Gilbert LI. 2006. The Halloween genes code for cytochrome P450 enzymes mediating synthesis of the insect moulting hormone. *Biochem Soc Trans* 34(Pt 6): 1256-60. PubMed ID: [17073797](#)
- Robinson GE, Hackett KJ, Purcell-Miramontes M, Brown SJ, Evans JD, Goldsmith MR, et al., Schneider DJ. 2011. Creating a buzz about insect genomes. *Science* 331(6023): 1386. PubMed ID: [21415334](#)
- Russo CA, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol* 12(3): 391-404. PubMed ID: [7739381](#)
- Scott JG. 1999. Cytochromes P450 and insecticide resistance. *Insect Biochem Mol Biol* 29(9): 757-77. PubMed ID: [10510498](#)
- Stegeman JJ, Livingstone DR. 1998. Forms and functions of cytochrome P450. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 121(1-3): 1-3. PubMed ID: [9972446](#)
- Stump AD, Jablonski SE, Bouton L, Wilder JA. 2011. Distribution and mechanism of α -amanitin tolerance in mycophagous *Drosophila* (Diptera: Drosophilidae). *Environ Entomol* 40(6): 1604-12. PubMed ID: [22217779](#)
- Sturtevant AH. (1916) Notes on North American Drosophilidae with descriptions of twenty-three new species. *Annals of the Entomological Society of America* 9(4): 323-343.
- Tamura K, Subramanian S, Kumar S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* 21(1): 36-44. PubMed ID: [12949132](#)
- Tello-Ruiz MK, Marco CF, Hsu FM, Khangura RS, Qiao P, Sapkota S, et al., Micklos DA. 2019. Double triage to identify poorly annotated genes in maize: The missing link in community curation. *PLoS One* 14(10): e0224086. PubMed ID: [31658277](#)
- Threlfall J, Blaxter M. 2021. Launching the Tree of Life Gateway. *Wellcome Open Res* 6: 125. PubMed ID: [34095514](#)
- Warren JT, Petryk A, Marques G, Jarcho M, Parvy JP, Dauphin-Villemant C, O'Connor MB, Gilbert LI. 2002. Molecular and biochemical characterization of two P450 enzymes in the ecdysteroidogenic pathway of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 99(17): 11043-8. PubMed ID: [12177427](#)
- Williams E, Chialvo P, Scott Chialvo C. 2026. Gene model for the ortholog of GstO3 in *Drosophila dunni*. *MicroPubl Biol* 2026: 10.17912/micropub.biology.002110. PubMed ID: [42294398](#)

Funding: This gene annotation project was funded by Nation Science Foundation grants DEB-1737869 (PI LKR, CoPI CSC) and DBI-2217912 (PI CSC). The Genomics Education Partnership (GEP; <https://thegep.org/>), which supports this project, is funded by the National Science Foundation (1915544; PI LKR) and the National Institute of General Medical Sciences of the National Institutes of Health (R25GM130517; PI LKR). Any opinions, findings, and conclusions or recommendations expressed in this material are solely those of the author(s) and do not necessarily reflect the official views of the National Science Foundation nor the National Institutes of Health.

Conflicts of Interest: The authors declare that there are no conflicts of interest present.

Author Contributions: Parker Brooks: data curation, formal analysis, investigation, writing - review editing. Pablo Chialvo: investigation, formal analysis, writing - review editing, validation. Clare Scott Chialvo: conceptualization, supervision, validation, writing - original draft.

Reviewed By: Geoff Findlay

Nomenclature Validated By: Anonymous

History: Received July 17, 2026 **Revision Received** July 27, 2026 **Accepted** July 30, 2026 **Published Online** July 31, 2026 **Indexed** August 14, 2026

Copyright: © 2026 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: Brooks P, Chialvo P, Scott Chialvo C. 2026. Gene model for the ortholog of *dib* in *Drosophila cardini*. *microPublication Biology*. [10.17912/micropub.biology.002318](#)

