# Gene model for the ortholog of *GstO3* in *Drosophila dunni*

Emma Williams[1], Pablo Chialvo[1], Clare Scott Chialvo[1§]
[1]Biology, App State, Boone, NC, US
[§]To whom correspondence should be addressed: chialvoch@appstate.edu

## Abstract

Developing a gene model for the *Glutathione S transferase O3* ortholog (*GstO3*) in the ASM1815212v1 Genome Assembly (GenBank Accession: GCA_018152125.1) of *Drosophila dunni*. 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.

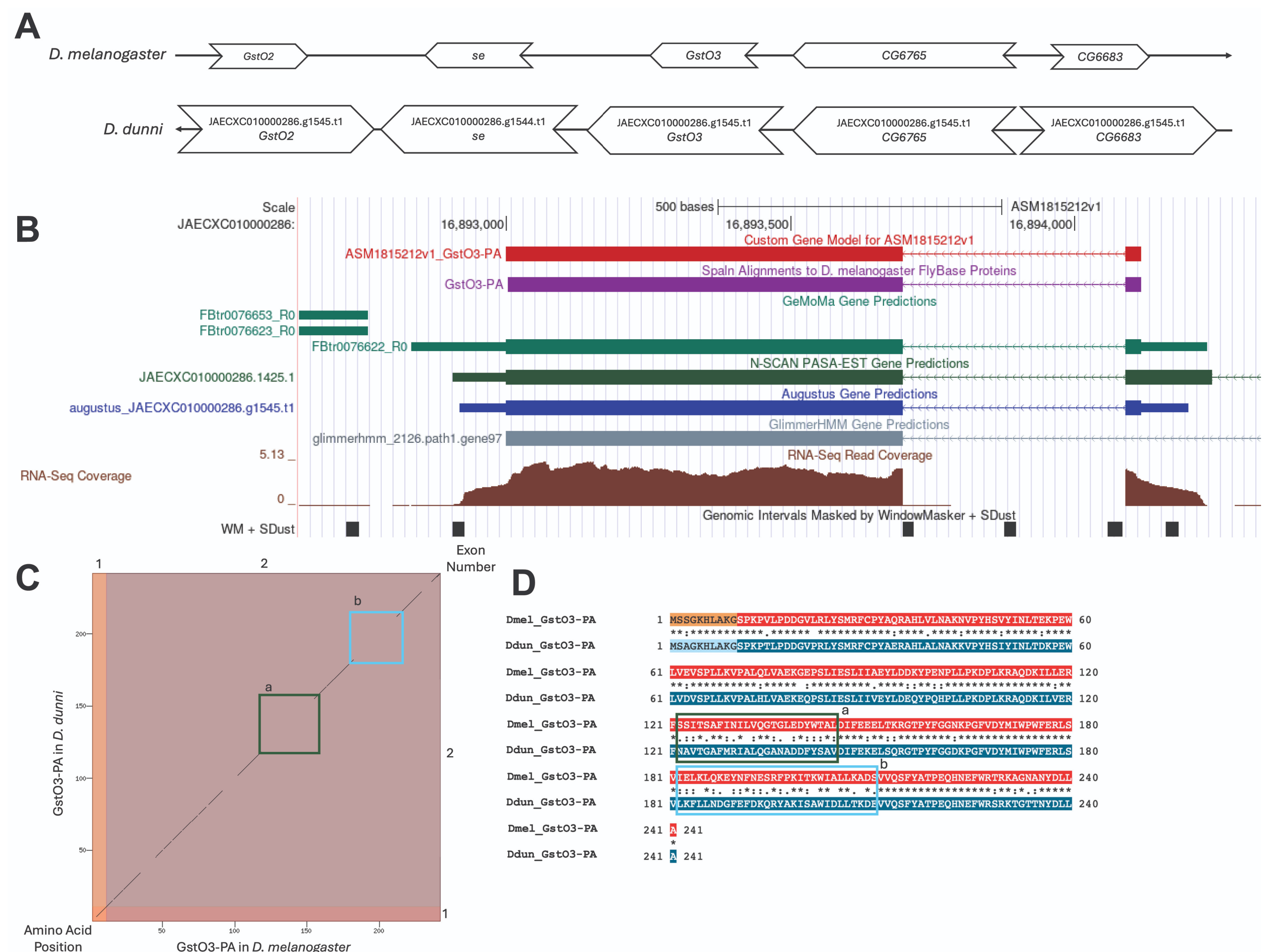


**Figure 1. Genomic neighborhood and gene model for *GstO3* in *D. dunni*:**

(A) **Synteny comparison of the genomic neighborhoods for *GstO3* in *Drosophila melanogaster* and *D. dunni*.** Thin underlying arrows indicate which DNA strand the target gene–*GstO3*–is located on in *D. melanogaster* (top) and *D. dunni* (bottom). The thin arrow pointing to the right indicates that *GstO3* is on the positive strand in *D. melanogaster*, and the thin arrow pointing to the left indicates that *GstO3* is on the negative strand in *D. dunni*. The wide gene arrows pointing in the same direction as *GstO3* are on the same strand relative to the thin underlying arrows, while wide gene arrows pointing in the opposite direction of *GstO3* are on the opposite strand relative to the thin underlying arrows. White gene arrows in *D. dunni* indicate orthology to the corresponding gene in *D. melanogaster*. Gene symbols given in the *D. dunni*gene arrows indicate the orthologous gene in *D. melanogaster*, while the gene prediction identifiers are specific to *D. dunni*. (B) **Gene Model in GEP UCSC Track Data Hub** (Raney et al., 2014). The coding-regions of *GstO3* in *D. dunni* 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, GeMoMa, and NSCAN PASA-EST (dark green), and RNA-Seq from adult flies of mixed sex (brown; alignment of Illumina RNA-Seq reads from *D. dunni* – Erlenbach et al. 2023). (C) **Dot Plot of GstO3-PA in *D. melanogaster* (*x*-axis) vs. the orthologous peptide in *D. dunni* (*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 with low sequence identity between species. In CDS 2, there are three small breaks and two longer breaks (dark green box – a; light blue box – b). (D) **Idiosyncrasies in protein alignment.** CDS 2 contains two longer breaks in the protein alignment that indicate low levels of sequence similarity. In the first long break (dark green box – a), over half of the 24 amino acids are conserved or highly chemically similar and only two are very dissimilar. The second long break (light blue box – b), covers 30 amino acids with 20 of them conserved or highly chemically similar and five that are very dissimilar. The three smaller breaks are associated with changes in one or two consecutive amino acids.

## Description

***Introduction -*** Within insects, the process of 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) is 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. dunni* for *Glutathione S transferase O3* (*GstO3*), a member of the glutathione S-transferase 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*.

Within the subgenus *Drosophila*, *D. dunni* Townsend and Wheeler 1955 is placed in the *dunni* subgroup of the *cardini* species group in the *immigrans-tripunctata* radiation (Heed & Krishnamurthy 1959; Bächli 2005). Species in the *dunni* subgroup, including *D. dunni*, are distributed across the Caribbean (Heed & Krishnamurthy 1959). Members of the *cardini* group primarily feed and develop on fruit and flowers (Markow & O'Grady 2008). While some mushroom-feeding *cardini* subgroup members tolerate the mushroom toxin a-amanitin (Stump et al. 2011), the fruit/flower feeding *dunni* subgroup species do not (Erlenbach et al. 2023).

One class of phase II detoxification enzymes are the glutathione S-transferases (GSTs), which act by conjugating xenobiotics or products of phase I detoxification with a glutathione to make them more hydrophilic prior to excretion (Enayati et al., 2005). GSTs are classified into two families based on their cellular location (microsomal and cytosolic). Within insects, cytosolic or canonical GSTs are divided into six classes (delta, sigma, epsilon, zeta, theta, and omega; Enayati et al., 2005; Ketterman et al., 2011). In the omega class, *Glutathione S transferase O3* (*GstO3*) is a biochemically active GST (Kim et al. 2006). It is significantly upregulated following exposure to oxidative stressors, such as hydrogen peroxide and zinc (Yepiskoposyan et al., 2006; Li et al., 2008). Across 12 *Drosophila* genomes, *GstO3* is highly conserved and expressed in all life stages of *D. melanogaster* (Walters et al., 2009).

We propose a gene model for the *D. dunni* ortholog of the *D. melanogaster Glutathione S transferase O3* (*GstO3*) gene. The genomic region of the ortholog corresponds to the Augustus gene prediction JAECXC010000286.g1545.t1 in the ASM1815212v1 Genome Assembly of *D. dunni* (GCA_018152125.1 – Kim et al., 2021). This model is based on mixed sex adult RNA-Seq data from *D. dunni* (Erlenbach et al. 2023; https://doi.org/10.5061/dryad.hdr7sqvq2) and *GstO3* in *D. melanogaster* using FlyBase release FB2024_02 (GCA_000001215.4; Gramates et al., 2022; Jenkins et al., 2022; Larkin et al., 2021).

***Synteny -*** The reference gene, *GstO3*, occurs on chromosome 3L in *D. melanogaster* and is flanked upstream by *CG6765* and *CG6683* and downstream by *sepia* (*se*) and *Glutathione S transferase O2* (*GstO2*). The *tblastn* search of *D. melanogaster* GstO3-PA (query) against the *D. dunni* (GenBank Accession: GCA_018152125.1 Genome Assembly (subject) placed the putative ortholog of *GstO3* in contig_291 (JAECXC010000286) which corresponds to Augustus gene prediction JAECXC010000286.g1545.t1 (E-value: 2e-139; percent identity: 74.27% as determined by *blastp*). The putative ortholog is flanked upstream by Augustus gene prediction JAECXC010000286.g1546.t1 and JAECXC010000286.g1547.t1, which correspond to *CG6765* and *CG6683* in *D. melanogaster* (E-value: 0.0 and 2e-52; identity: 63.37% and 49.54%, respectively, as determined by *blastp*; Figure 1A; Altschul et al., 1990). The putative ortholog of *GstO3* is flanked downstream by Augustus gene prediction JAECXC010000286.g1544.t1 and JAECXC010000286.g1543.t1, which correspond to *se* and *GstO2* in *D. melanogaster* (E-value: 5e-167 and 2e-127;

identity: 90.12% and 68.80%, respectively, as determined by *blastp*). The putative ortholog assignment for *GstO3* in *D. dunni* is supported by the following evidence: the gene predictions surrounding the *GstO3* 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 Augustus gene prediction JAECXC010000286.g1545.t1 is an ortholog of *GstO3* in *D. dunni* (Figure 1A).

***Protein Model*** - *GstO3* in *D. dunni* has 2 coding sequences (CDS) within the genome sequence. The only unique protein sequence (GstO3-PA) is translated from 1 mRNA isoform (GstO3-RA; Figure 1B). Relative to the ortholog in *D. melanogaster*, the CDS number and protein isoform count are conserved (2 CDSs and 1 mRNA isoform*)*. The sequence ofGstO3-PA in *D. dunni* has 74.3% identity (88.0% similarity) with the protein-coding isoform GstO3-PA in *D. melanogaster*, as determined by *blastp* (Figure 1C). This level of divergence is not surprising given that *D. dunni* and *D. melanogaster* belong to two separate subgenera (*Drosophila* and *Sophophora* respectively) that diverged approximately 45-60MYA (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. dunni* (GCA_018152125.1 – 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 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 sex 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. melanogaste*r 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. FlyBase is supported by grants: NHGRI U41HG000739 and U24HG010859, UK Medical Research Council MR/W024233/1, NSF 2035515 and 2039324, BBSRC BB/T014008/1, and Wellcome Trust PLM13398.

## Extended Data

Description: Zipped archive containing FASTA, PEP, and GFF files. Resource Type: “Model” . Resource Type: Dataset. File: Ddun_GstO3_Model.tar.gz. DOI: 10.22002/p58qq-wbm15

## 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 August 2024, last accessed 19 September 2024. http://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

Enayati AA, Ranson H, Hemingway J. 2005. Insect glutathione transferases and insecticide resistance. Insect Mol Biol 14(1): 3-8. PubMed ID: 15663770

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

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.

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

Ketterman AJ, Saisawang C, Wongsantichon J. 2011. Insect glutathione transferases. Drug Metab Rev 43(2): 253-65. PubMed ID: 21323601

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

Kim J, Suh H, Kim S, Kim K, Ahn C, Yim J. 2006. Identification and characteristics of the structural gene for the Drosophila eye colour mutant sepia, encoding PDA synthase, a member of the omega class glutathione S-transferases. Biochem J 398(3): 451-60. PubMed ID: 16712527

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 HM, Buczkowski G, Mittapalli O, Xie J, Wu J, Westerman R, et al., Pittendrigh BR. 2008. Transcriptomic profiles of Drosophila melanogaster third instar larval midgut and responses to oxidative stress. Insect Mol Biol 17(4): 325-39. PubMed ID: 18651915

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

Rane RV, Walsh TK, Pearce SL, Jermiin 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

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

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

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

Townsend JI, Wheeler MR. (1955) Notes on Puerto Rican Drosophilidae, including descriptions of two new species of *Drosophila*. *The Journal of Agriculture of the University of Puerto Rico* 39(2): 57-64.

Walters KB, Grant P, Johnson DL. 2009. Evolution of the GST omega gene family in 12 Drosophila species. J Hered 100(6): 742-53. PubMed ID: 19608790

Yepiskoposyan H, Egli D, Fergestad T, Selvaraj A, Treiber C, Multhaup G, Georgiev O, Schaffner W. 2006. Transcriptome response to heavy metal stress in Drosophila reveals a new zinc transporter that confers resistance to zinc. Nucleic Acids Res 34(17): 4866-77. PubMed ID: 16973896

**Funding:** This gene annotation project was funded by Nation Science Foundation grants DEB-1737869 (PI L.K. Reed, CoPI C. Scott Chialvo) and DBI-2217912 (PI C. Scott Chialvo). The Genomics Education Partnership (GEP; https://thegep.org/), which supports this project, is funded by the National Science Foundation (1915544; PI L.K. Reed) and the National Institute of General Medical Sciences of the National Institutes of Health (R25GM130517; PI L.K. Reed). 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:** Emma Williams: data curation, formal analysis, writing - review editing, investigation. Pablo Chialvo: investigation, formal analysis, writing - review editing, validation. Clare Scott Chialvo: conceptualization, supervision, validation, writing - original draft.

**Reviewed By:** Anonymous

**Nomenclature Validated By:** Anonymous

**History: Received** March 20, 2026 **Revision Received** April 8, 2026 **Accepted** May 26, 2026 **Published Online** May 27, 2026 **Indexed** June 10, 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:** Williams E, Chialvo P, Scott Chialvo C. 2026. Gene model for the ortholog of *GstO3* in *Drosophila dunni*. microPublication Biology. 10.17912/micropub.biology.002110