microPublication BIOLOGY

9/29/2026 - *Open Access*

Gene model for the ortholog of *snu* in *Drosophila funebris*

Braiden S. McAlpin[1], Pablo Chialvo[1], Clare Scott Chialvo[1§]

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

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

Abstract

We developed a gene model for the ortholog of the *Drosophila melanogaster* gene *snustorr* (*snu*) in the ASM1890182v1 Genome Assembly (GenBank Accession: GCA_018901825.1) of *Drosophila funebris*. 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.

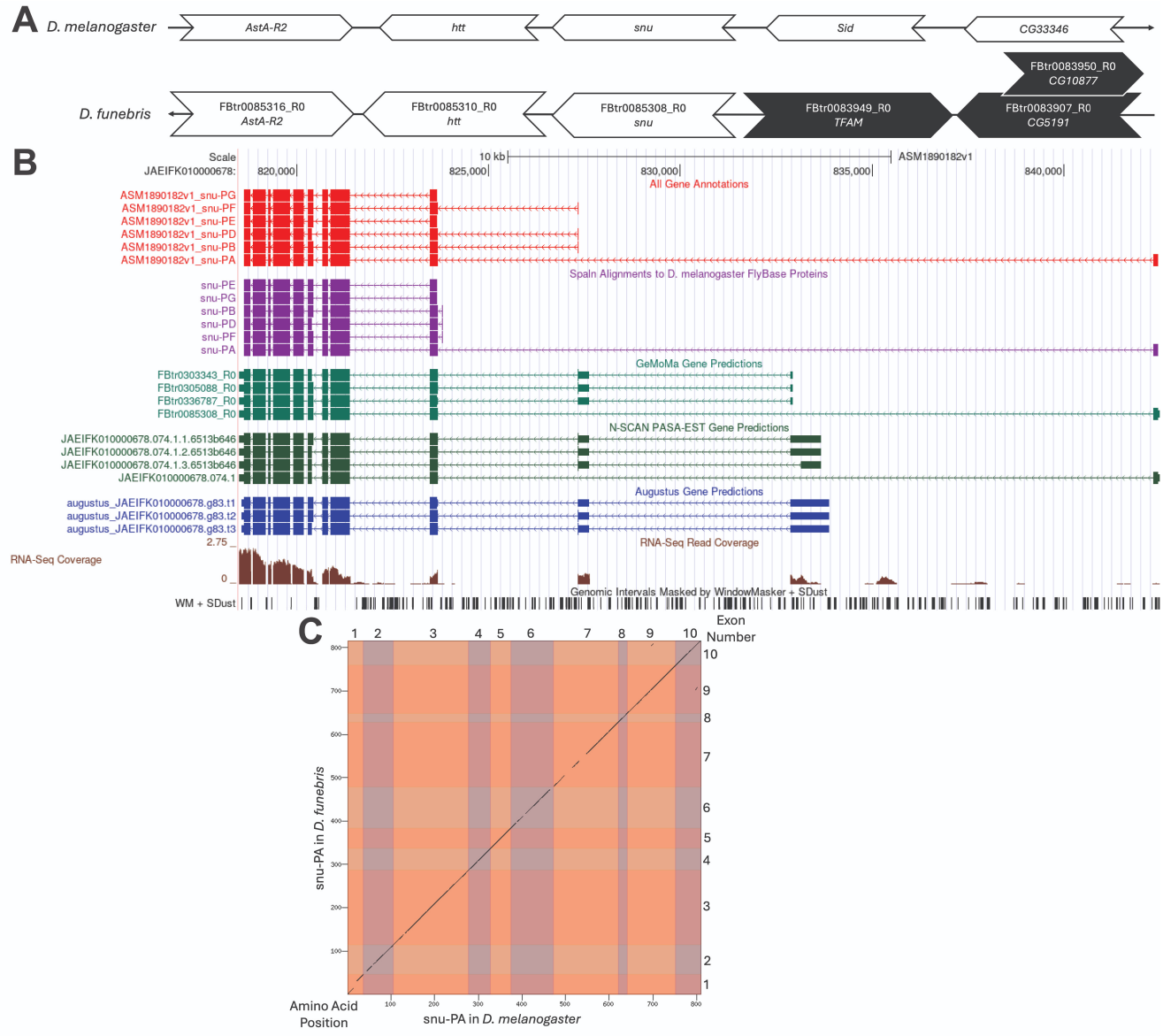

Figure 1. Genomic neighborhood and gene model for *snu* ortholog in *D. funebris*:

(A) **Synteny comparison of the genomic neighborhoods for *snu* in *Drosophila melanogaster* and *Drosophila funebris*.** Thin underlying arrows indicate which DNA strand the target gene, *snu*, is located on in *D. melanogaster* (top) and *D. funebris* (bottom). The thin arrow pointing to the right indicates that *snu* is on the positive strand in *D. melanogaster*, and the thin arrow pointing to the left indicates that *snu* is on the negative strand in *D. funebris*. The wide gene arrows pointing in the same direction as *snu* are on the same strand relative to the thin underlying arrows, while wide gene arrows

pointing in the opposite direction of *snu* are on the opposite strand relative to the thin underlying arrows. White gene arrows in *D. funebris* indicate orthology to the corresponding gene in *D. melanogaster*, while black gene arrows indicate non-orthology. Gene symbols given in the *D. funebris* gene arrows indicate the orthologous gene in *D. melanogaster*, while the locus identifiers are specific to *D. funebris*. (B) **Gene Model in GEP UCSC Track Data Hub** (Raney et al., 2014). The coding-regions of *snu* in *D. funebris* 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. funebris* – Erlenbach et al., 2023). (C) **Dot Plot of snu-PA in *D. melanogaster* (*x*-axis) vs. the orthologous peptide in *D. funebris* (*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.

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. funebris* for *snustorr* (*snu*), a member of the ABC transporter 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)

"*Drosophila funebris* (Fabricius, 1787) is a member of the *funebris* species group, which occurs in the *immigrans-tripunctata* radiation of the *Drosophila* subgenus (Bächli, 2005; ICZN, 2010). It is also the type species of the genus *Drosophila*. This species is a globally distributed human commensal (Grimaldi, 2022). While *D. funebris* feeds on shelf fungi and fruit (Kimura et al., 1977; Prigent et al., 2003), it does not tolerate the mushroom toxin α-amanitin (Stump et al., 2011; Erlenbach et al., 2023)." (McDonald et al., 2026).

"ATP-binding cassette (ABC) transporters constitute one of the largest protein superfamilies across all domains of life, functioning as primary-active transporters that use the energy of ATP hydrolysis to move a broad range of substrates across lipid membranes (Higgins, 1992; Dean et al., 2001). In the context of xenobiotic metabolism, ABC transporters mediate phase III detoxification by excreting toxins along with phase I and phase II metabolites out of cells. Their overexpression is a recurrent mechanism of insecticide and multidrug resistance in insects (Merzendorfer, 2014; Dermauw and Van Leeuwen, 2014)." (Dieterle et al., 2026)

snustorr (*snu*) is a member of the ABCH ATP-binding cassette transporter subfamily of ABC transporters. It is required for cuticle development and localizes to the apical plasma membrane and cytoplasmic vesicles, consistent with a role in exporting lipids to build the epidermal barrier (Zuber et al., 2018; Wang et al., 2020). Because the cuticular barrier is the insect's first defense against xenobiotic penetration, *snu* is considered part of the "Phase 0" detoxification program (Gao et al., 2022). Transcripts are enriched in the embryonic tracheal system and epidermis (Öztürk-Çolak et al., 2024).

We propose a gene model for the *D. funebris* ortholog of the *D. melanogaster snustorr* (*snu*) gene. The genomic region of the ortholog corresponds to the GeMoMa FBtr0085308_R0 gene prediction in the ASM1890182v1 Genome Assembly of *D. funebris* (GCA_018901825.1 – Kim et al, 2021). This model is based on mixed sex, adult RNA-Seq data from *D.*

funebris (Erlenbach et al., 2023; https://doi.org/10.5061/dryad.hdr7sqvq2) and *snu* in *D. melanogaster* using FlyBase release FB2024_02 (GCA_000001215.4; Öztürk-Çolak et al., 2024).

Synteny

The reference gene, *snu*, occurs on chromosome 3R in *D. melanogaster* and is flanked upstream by *huntingtin* (*htt*) and *Allatostatin A receptor 2* (*AstA-R2*) and downstream by *Stress induced DNase* (*Sid*) and *CG33346*. The *tblastn* search of *D. melanogaster* snu-PA (query) against the *D. funebris* (GenBank Accession: GCA_018901825.1) Genome Assembly (subject) placed the putative ortholog of *snu* within contig_683 (JAEIFK010000678) which corresponds to GeMoMa gene prediction FBtr0085308_R0 (E-value: 0; percent identity: 92.53% as determined by *blastp*). The putative ortholog is flanked downstream by FBtr0083949_R0, FBtr0083907_R0, and FBtr0083950_R0 (nested) which correspond to *mitochondrial transcription factor A* (*TFAM*), *CG5191*, and *CG10877* in *D. melanogaster* (E-value: 7e-142, 0, 0; identity: 69.42%, 96.97%, 70.79% respectively, as determined by *blastp*; Figure 1A; Altschul et al., 1990). The putative ortholog of *snu* is flanked upstream by FBtr0085310_R0 and FBtr0085316_R0, which correspond to *htt* and *AstA-R2* in *D. melanogaster* (E-value: 0 and 6e-179; identity:66.34% and 70.95%, respectively, as determined by *blastp*). The putative ortholog assignment for *snu* in *D. funebris* is supported by the following evidence: the *tblastn* results are of good quality, and all coding sequences (CDS) and isoforms found in *D. melanogaster* also appear to be present in *D. funebris*. Though the gene predictions immediately downstream of the putative *snu* ortholog are not conserved, they do appear in the same order and orientation much further downstream in *D. melanogaster*. Conversely, the upstream genetic neighborhood of the putative ortholog is conserved. As such, we conclude that FBtr0085308_R0 is an ortholog of *snu* in *D. funebris* (Figure 1A).

Protein Model

snu in *D. funebris* has ten CDS within the genomic sequence. All six isoforms (PA-PE, Figure 1B) are translated from unique messenger RNAs. Relative to the ortholog in *D. melanogaster*, the CDS number and protein isoform count are conserved. The sequence of snu-PA in *D. funebris* has 92.5% identity (96.1% similarity) with the protein-coding isoform snu-PA in *D. melanogaster*, as determined by *blastp* (Figure 1C). 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 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. funebris* (GCA_018901825.1 – 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 mixed sex RNA-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 of snu model in D. funebris. Resource Type: Model. File: Dfun_snu_Model.tar.gz. DOI: 10.22002/tqrd0-6z073

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/

Dean M, Rzhetsky A, Allikmets R. 2001. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res 11(7): 1156-66. PubMed ID: 11435397

Dermauw W, Van Leeuwen T. 2014. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol 45: 89-110. PubMed ID: 24291285

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

Dieterle S, Chialvo P, Scott Chialvo C. 2026. Gene model for the ortholog of *ABCB7* in *Drosophila funebris*. microPublication Biology. 10.17912/micropub.biology.002346. DOI: 10.17912/micropub.biology.002346

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

Fabricius JC, Metcalf Collection (North Carolina State University)., Tippmann Collection (North Carolina State University).. 1787. Mantissa insectorum sistens eorum species nuper detectas adiectis characteribus genericis, differentiis specificis, emendationibus, observationibus, Ioh. Christ. Fabricii.. : 10.5962/bhl.title.36471. DOI: 10.5962/bhl.title.36471

Gao L, Zang X, Qiao H, Moussian B, Wang Y. 2022. Xenobiotic responses of Drosophila melanogaster to insecticides with different modes of action and entry. Arch Insect Biochem Physiol 111(4): e21958. PubMed ID: 35942563

Grimaldi DA. 2022. The Drosophila funebris Species Group in North America (Diptera: Drosophilidae). American Museum Novitates 2022: 10.1206/3988.1. DOI: 10.1206/3988.1

Higgins CF. 1992. ABC transporters: from microorganisms to man. Annu Rev Cell Biol 8: 67-113. PubMed ID: 1282354

ICZN. (2010) Opinion 2245 (Case 3407): *Drosophila* Fallén, 1823 (Insecta, Diptera): *Drosophila funebris* Fabricius, 1787 is maintained as the type species. *The Bulletin of Zoological Nomenclature* 67(1), 106-115.

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

Kimura M, Toda M, Beppu K, Watabe H. (1977) Breeding sites of Drosophilid flies in and near Sapporo northern Japan, with supplementary notes on adult feeding habits. *Japanese Journal of Entomology* 45: 571-582.

McDonald M, Chialvo P, Scott Chialvo C. 2026. Gene model for the ortholog of *PGRP-SB1* in *Drosophila funebris*. microPublication Biology. 10.17912/micropub.biology.002270.

Merzendorfer H. 2014. ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites. Advances in Insect Physiology,Target Receptors in the Control of Insect Pests: Part II : 1-72. DOI: 10.1016/B978-0-12-417010-0.00001-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

Öztürk-Çolak A, Marygold SJ, Antonazzo G, Attrill H, Goutte-Gattat D, Jenkins VK, et al., FlyBase Consortium. 2024. FlyBase: updates to the Drosophila genes and genomes database. Genetics 227(1): 10.1093/genetics/iyad211. PubMed ID: 38301657

Prigent Sp, Renard E, Cariou ML. 2003. Electrophoretic Mobility of Amylase in Drosophilids Indicates Adaptation to Ecological Diversity. Genetica 119: 133-145. DOI: 10.1023/A:1026071317995

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

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

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

Wang Y, Norum M, Oehl K, Yang Y, Zuber R, Yang J, et al., Moussian B. 2020. Dysfunction of Oskyddad causes Harlequin-type ichthyosis-like defects in Drosophila melanogaster. PLoS Genet 16(1): e1008363. PubMed ID: 31929524

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

Zuber R, Norum M, Wang Y, Oehl K, Gehring N, Accardi D, et al., Moussian B. 2018. The ABC transporter Snu and the extracellular protein Snsl cooperate in the formation of the lipid-based inward and outward barrier in the skin of Drosophila. Eur J Cell Biol 97(2): 90-101. PubMed ID: 29306642

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: Braiden S. McAlpin: data curation, formal analysis, investigation, writing - review editing. Pablo Chialvo: investigation, formal analysis, validation, writing - original draft. Clare Scott Chialvo: conceptualization, supervision, validation, writing - review editing.

Reviewed By: Anonymous, Brandon Turner, Cedric Aumont

Nomenclature Validated By: Anonymous

History: Received August 12, 2026 **Revision Received** September 21, 2026 **Accepted** September 29, 2026 **Published Online** September 29, 2026 **Indexed** October 13, 2026

Citation: McAlpin BS, Chialvo P, Scott Chialvo C. 2026. Gene model for the ortholog of *snu* in *Drosophila funebris*. microPublication Biology. 10.17912/micropub.biology.002347