

Gene model for the ortholog of *Glys* in *Drosophila simulans*

Madeline L. Gruys¹, Madison A. Sharp¹, Zachary Lill², Caroline Xiong³, Amy T. Hark⁴, James J. Youngblom³, Chinmay P. Rele¹, Laura K Reed^{1§}

¹The University of Alabama, Tuscaloosa, AL USA

²Muhlenburg College, Allentown, PA USA

³California State University Stanislaus, Turlock, CA USA

⁴Muhlenburg College, Allentown, PA USA

[§]To whom correspondence should be addressed: lreed1@ua.edu

Abstract

Gene model for the ortholog of glycogen synthase (*Glys*) in the *Drosophila simulans* May 2017 (Princeton ASM75419v2/DsimGB2) Genome Assembly (GenBank Accession: [GCA_000754195.3](https://www.ncbi.nlm.nih.gov/assembly/GCA_000754195.3)). This ortholog was characterized as part of a developing dataset to study the evolution of the Insulin/insulin-like growth factor signaling pathway (IIS) across the genus *Drosophila* using the Genomics Education Partnership gene annotation protocol for Course-based Undergraduate Research Experiences.

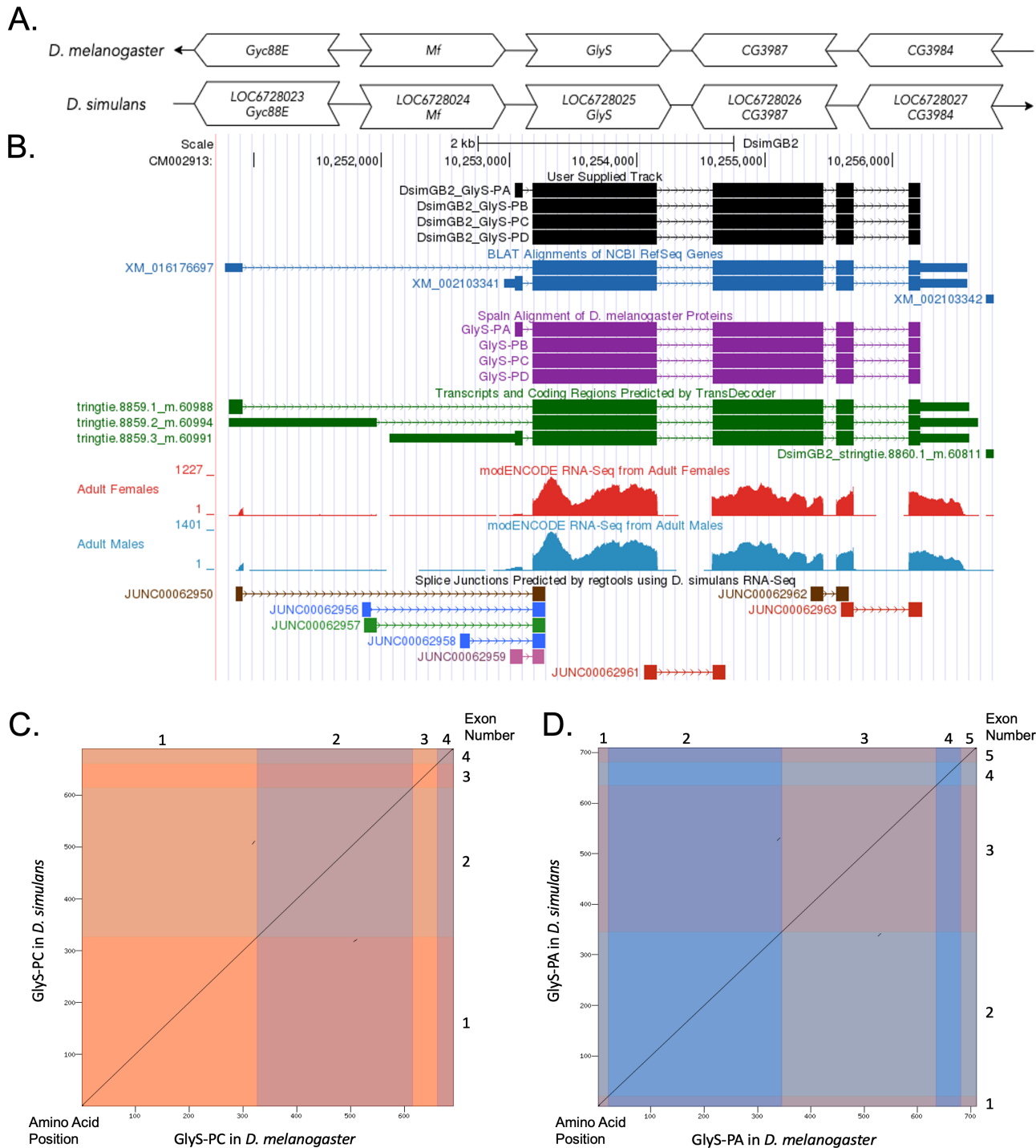


Figure 1. GlyS gene model comparison between *Drosophila simulans* and *Drosophila melanogaster* orthologs:

(A) Synteny comparison of the genomic neighborhoods for *GlyS* in *Drosophila melanogaster* and *D. simulans*. This underlying arrows indicate the DNA strand within which the target gene—*GlyS*—is located in *D. melanogaster* (top) and *D. simulans* (bottom) genomes. The thin arrow pointing to the right indicates that *GlyS* is on the positive (+) strand in *D. simulans*, and the thin arrow pointing to the left indicates that *GlyS* is on the negative (-) strand in *D. melanogaster*. The wide gene arrows pointing in the same direction as *GlyS* are on the same strand relative to the thin underlying arrows, while wide gene arrows pointing in the opposite direction of *GlyS* are on the opposite strand relative to the thin underlying arrows. White gene arrows in *D. simulans* indicate orthology to the corresponding gene in *D. melanogaster*. Gene symbols given in the *D. simulans* gene arrows indicate the orthologous gene in *D. melanogaster*, while the locus identifiers are specific to *D. simulans*.

(B) Gene Model in GEP UCSC Track Data Hub (Raney et al., 2014). The coding-regions of *GlyS* in *D. simulans* are

displayed in the User Supplied Track (black); CDSs are depicted by thick rectangles and introns by thin lines with arrows indicating the direction of transcription. Subsequent evidence tracks include BLAT Alignments of NCBI RefSeq Genes (dark blue, alignment of Ref-Seq genes for *D. simulans*), Spaln of *D. melanogaster* Proteins (purple, alignment of Ref-Seq proteins from *D. melanogaster*), Transcripts and Coding Regions Predicted by TransDecoder (dark green), RNA-Seq from Adult Females and Adult Males (red and light blue, respectively; alignment of Illumina RNA-Seq reads from *D. simulans*), and Splice Junctions Predicted by regtools using *D. simulans* RNA-Seq (SRP006203). Splice junctions shown have a read-depth of 10-49, 50-99, 100-499, 500-999, >1000 supporting reads in blue, green, pink, brown, and red, respectively. **(C) Dot Plot of GlyS-PC in *D. melanogaster* (x-axis) vs. the orthologous peptide in *D. simulans* (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. **(D) Dot Plot of GlyS-PA in *D. melanogaster* (x-axis) vs. the orthologous peptide in *D. simulans* (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.

Description

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 this box may be repeated in other articles submitted by participants using the same GEP CURE protocol for annotating Drosophila species orthologs of Drosophila melanogaster genes in the insulin signaling pathway.

"In this GEP CURE protocol students use web-based tools to manually annotate genes in non-model *Drosophila* species based on orthology to genes in the well-annotated model organism fruitfly *Drosophila melanogaster*. The GEP uses web-based tools to allow undergraduates to participate in course-based research by generating manual annotations of genes in non-model species (Rele et al., 2023). Computational-based gene predictions in any organism are often improved by careful manual annotation and curation, allowing for more accurate analyses of gene and genome evolution (Mudge and Harrow 2016; Tello-Ruiz et al., 2019). These models of orthologous genes across species, such as the one presented here, then provide a reliable basis for further evolutionary genomic analyses when made available to the scientific community." (Myers et al., 2024).

"The particular gene ortholog described here was characterized as part of a developing dataset to study the evolution of the Insulin/insulin-like growth factor signaling pathway (IIS) across the genus *Drosophila*. The Insulin/insulin-like growth factor signaling pathway (IIS) is a highly conserved signaling pathway in animals and is central to mediating organismal responses to nutrients (Hietakangas and Cohen 2009; Grewal 2009)." (Myers et al., 2024).

"*D. simulans* is part of the *melanogaster* species group within the subgenus *Sophophora* of the genus *Drosophila* (Sturtevant 1939; Bock and Wheeler 1972). It was first described by Sturtevant (1919). *D. simulans* is a sibling species to *D. melanogaster*, thus extensively studied in the context of speciation genetics and evolutionary ecology (Powell 1997). Historically, *D. simulans* was a tropical species native to sub-Saharan Africa (Lemeunier et al., 1986) where figs served as a primary host (Lachaise and Tsacas 1983). However, *D. simulans*'s range has expanded worldwide within the last century as a human commensal using a broad range of rotting fruits as breeding sites (<https://www.taxodros.uzh.ch>, accessed 1 Feb 2023)." (Lawson et al., 2024).

"*Glycogen synthase* (*Glys*; aka. *GS*, *GlyS*) is a gene within the Insulin-signaling pathway in *Drosophila* and encodes a glycosyltransferase that catalyzes linkage of glucose monomers into glycogen. *Glys* activity is regulated allosterically by glucose 6-phosphate and phosphorylation/dephosphorylation allowing for control of cellular glycogen levels (Plyte et al., 1992; Roach et al., 2012). Null *Glys* mutants exhibit growth defects and reduced larval viability in *Drosophila* (Yamada et al., 2019)." (Backlund et al., 2024).

We propose a gene model for the *D. simulans* ortholog of the *D. melanogaster* Glycogen synthase (*Glys*) gene. The genomic region of the ortholog corresponds to the uncharacterized protein [LOC6728025](#) (RefSeq accession [XP_016034640.1](#)) in the ASM75419v3 Genome Assembly of *D. simulans* (GenBank Accession: [GCA_000754195.3](#) - Hu et al., 2013). This model is based on RNA-Seq data from *D. simulans* ([SRP006203](#) - Graveley et al., 2011) and *Glys* in *D. melanogaster* using FlyBase release FB2022_04 ([GCA_000001215.4](#); Larkin et al., 2021). Gene and species details can be found in the description above.

Synteny

The reference gene, *Glys*, occurs on chromosome 3R in *D. melanogaster* and is flanked upstream by *Guanylyl cyclase at 88E* (*Gyc88E*) and *Myofilin* (*Mf*) and downstream by *CG3987* and *CG3984*. The *tblastn* search of *D. melanogaster* *Glys*-PC (query) against the *D. simulans* (GenBank Accession: [GCA_000754195.3](#)) Genome Assembly (database) placed the putative ortholog of *Glys* within scaffold CM002913 (CM002913.1) at locus [LOC6728025](#) ([XP_016034640.1](#))— with an E-value of 0.0 and a percent identity of 100.00%. Furthermore, the putative ortholog is flanked upstream by [LOC6728023](#) ([XP_016034621.1](#)) and [LOC6728024](#) ([XP_044779336.1](#)), which correspond to *Gyc88E* and *Mf* in *D. melanogaster* (E-value: 0.0 and 0.0; identity: 97.27% and 97.82%, respectively, as determined by *blastp*; Figure 1A, Altschul et al., 1990). The putative ortholog of *Glys* is flanked downstream by [LOC6728026](#) ([XP_002103378.1](#)) and [LOC6728027](#) ([XP_016034642.1](#)), which correspond to *CG3987* and *CG3984* in *D. melanogaster* (E-value: 0.0 and 6e-168; identity: 87.14% and 85.13%, respectively, as determined by *blastp*). The putative ortholog assignment for *Glys* in *D. simulans* is supported by the following evidence: The genes surrounding the *Glys* ortholog are orthologous to the genes at the same locus in *D. melanogaster* and local synteny is completely conserved, supported by results generated from *blastp*, so we conclude that [LOC6728025](#) is the correct ortholog of *Glys* in *D. simulans* (Figure 1A).

Protein Model

Glys in *D. simulans* has two unique protein-coding isoforms (*Glys*-PA and *Glys*-PC, identical to *Glys*-PB and *Glys*-PD; Figure 1B). mRNA isoforms *Glys*-RC, *Glys*-RB, and *Glys*-RD, which differ in their UTRs, contain four CDSs, and isoform *Glys*-RA contains five CDSs. Relative to the ortholog in *D. melanogaster*, the RNA CDS number is conserved. The sequence of *Glys*-PC in *D. simulans* has 100.00% identity (E-value: 0.0) with the protein-coding isoform *Glys*-PC in *D. melanogaster*, as determined by *blastp* (Figure 1C). The dot plot of the ortholog with *Glys*-PA is exhibited in Figure 1D (E-value: 0.0; identity: 100.00%, as determined by *blastp*). Coordinates of this curated gene model (*Glys*-PA, *Glys*-PB, *Glys*-PC, *Glys*-PD) are stored by NCBI at GenBank/BankIt (accession [BK064593](#), [BK064594](#), [BK064595](#), [BK064596](#), respectively). These data are also archived in the CaltechDATA repository (see “Extended Data” section below).

Methods

Detailed methods including algorithms, database versions, and citations for the complete annotation process can be found in Rele et al. (2023). Briefly, students use the GEP instance of the UCSC Genome Browser v.435 (<https://gander.wustl.edu>; Kent WJ et al., 2002; Navarro Gonzalez et al., 2021) to examine the genomic neighborhood of their reference IIS gene in the *D. melanogaster* genome assembly (Aug. 2014; BDGP Release 6 + ISO1 MT/dm6). Students then retrieve the protein sequence for the *D. melanogaster* reference gene for a given isoform and run it using *tblastn* against their target *Drosophila* species genome assembly on the NCBI BLAST server (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>; Altschul et al., 1990) to identify potential orthologs. To validate the potential ortholog, students compare the local genomic neighborhood of their potential ortholog with the genomic neighborhood of their reference gene in *D. melanogaster*. This local synteny analysis includes at minimum the two upstream and downstream genes relative to their putative ortholog. They also explore other sets of genomic evidence using multiple alignment tracks in the Genome Browser, including BLAT alignments of RefSeq Genes, Spaln alignment of *D. melanogaster* proteins, multiple gene prediction tracks (e.g., GeMoMa, Geneid, Augustus), and modENCODE RNA-Seq from the target species. Detailed explanation of how these lines of genomic evidenced are leveraged by students in gene model development are described in Rele et al. (2023). Genomic structure information (e.g., CDSs, intron-exon number and boundaries, number of isoforms) for the *D. melanogaster* reference gene is retrieved through the Gene Record Finder (<https://gander.wustl.edu/~wilson/dmelgenerecord/index.html>; Rele et al., 2023). Approximate splice sites within the target gene are determined using *tblastn* using the CDSs from the *D. melanogaster* reference gene. Coordinates of CDSs are then refined by examining aligned modENCODE RNA-Seq data, and by applying paradigms of molecular biology such as identifying canonical splice site sequences and ensuring the maintenance of an open reading frame across hypothesized splice sites. Students then confirm the biological validity of their target gene model using the Gene Model Checker (<https://gander.wustl.edu/~wilson/dmelgenerecord/index.html>; Rele et al., 2023), which compares the structure and translated sequence from their hypothesized target gene model against the *D. melanogaster* reference gene model. At least two independent models for a gene are generated by students under mentorship of their faculty course instructors. Those models are then reconciled by a third independent researcher mentored by the project leaders to produce the final model. Note: comparison of 5' and 3' UTR sequence information is not included in this GEP CURE protocol.

Acknowledgements: This publication is dedicated to the memory of Dr. James J. Youngblom. We would like to thank Wilson Leung for developing and maintaining the technological infrastructure that was used to create this gene model, Cole A. Kiser and Leon F. Laskowski for their help reviewing the model. 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: File contains the GFF, amino acid, and nucleic acid sequences for the gene model.. Resource Type: Model. File: [DsimGB2_Glys.zip](#). DOI: [10.22002/ywzts-78198](#)

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](#)
- Backlund AE, White J, Grillo L, Ianniello T, Swedrowski A, Yang M, Jemc J, Rele CP, 2024. Gene model for the ortholog of *Glys* in *Drosophila busckii*. *microPublication Biology*, submitted PubMed ID: [null](#)
- Bock IR, Wheeler MR. 1972. The *Drosophila melanogaster* species-group. The University of Texas Publication. 7213: 1-102. FBrf0024428. PubMed ID: [null](#)
- Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al., Celniker SE. 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471(7339): 473-9. PubMed ID: [21179090](#)
- Grewal SS. 2009. Insulin/TOR signaling in growth and homeostasis: a view from the fly world. *Int J Biochem Cell Biol* 41(5): 1006-10. PubMed ID: [18992839](#)
- Hietakangas V, Cohen SM. 2009. Regulation of tissue growth through nutrient sensing. *Annu Rev Genet* 43: 389-410. PubMed ID: [19694515](#)
- Hu TT, Eisen MB, Thornton KR, Andolfatto P. 2013. A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. *Genome Res* 23(1): 89-98. PubMed ID: [22936249](#)
- 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](#)
- Koehler AC, Romo I, Le V, Youngblom JJ, Hark AT, Rele CP, Reed LK. 2024. Gene model for the ortholog of *Glys* in *Drosophila yakuba*. *MicroPubl Biol* 2024. PubMed ID: [null](#)
- Lachaise D, Tsacas L. (1983). Breeding-sites of tropical African *Drosophilids*. Ashburner, Carson, Thompson, 1981-1986 3d: 221-332. FBrf0038884 PubMed ID: [null](#)
- 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](#)
- Lawson ME, Jones GM, Runion M, Sims D, Young A, Briggs O, et al., Reed LK. 2024. Gene model for the ortholog of *ImpL2* in *Drosophila simulans*. *MicroPubl Biol* 2024. PubMed ID: [39677518](#)
- Lemeunier F, David J, Tsacas L, Ashburner M. 1986. The *melanogaster* species group. Ashburner, Carson, Thompson, 1981-1986 e: 147-256. FBrf0043749 PubMed ID: [null](#)
- Mudge JM, Harrow J. 2016. The state of play in higher eukaryote gene annotation. *Nat Rev Genet* 17(12): 758-772. PubMed ID: [27773922](#)
- Myers A, Hoffman A, Natysin M, Arsham AM, Stamm J, Thompson JS, Rele CP, Reed LK. 2024. Gene model for the ortholog *Myc* in *Drosophila ananassae*. *MicroPubl Biol* 2024. PubMed ID: [39677519](#)
- Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, et al., Kent WJ. 2021. The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res* 49(D1): D1046-D1057. PubMed ID: [33221922](#)
- Plyte SE, Hughes K, Nikolakaki E, Pulverer BJ, Woodgett JR. 1992. Glycogen synthase kinase-3: functions in oncogenesis and development. *Biochim Biophys Acta* 1114(2-3): 147-62. PubMed ID: [1333807](#)
- Powell JR. 1997. Progress and prospects in evolutionary biology: the *Drosophila* model. Oxford University Press, ISBN: 9780195076912 PubMed ID: [null](#)
- 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](#)
- Rele CP, Sandlin KM, Leung W, Reed LK. (2020). Manual Annotation of Genes within *Drosophila* Species: the Genomics Education Partnership protocol. *bioRxiv* 2020.12.10.420521 PubMed ID: [null](#)

Roach PJ, Depaoli-Roach AA, Hurley TD, Tagliabracci VS. 2012. Glycogen and its metabolism: some new developments and old themes. *Biochem J* 441(3): 763-87. PubMed ID: [22248338](#)

Sturtevant AH, 1919. A new species closely resembling *Drosophila melanogaster*. *Psyche* 26: 488-500. FBrf0000977 PubMed ID: [null](#)

Sturtevant AH. 1939. On the Subdivision of the Genus *Drosophila*. *Proc Natl Acad Sci U S A* 25(3): 137-41. PubMed ID: [16577879](#)

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](#)

Yamada T, Habara O, Yoshii Y, Matsushita R, Kubo H, Nojima Y, Nishimura T. 2019. The role of glycogen in development and adult fitness in *Drosophila*. *Development* 146(8). PubMed ID: [30918052](#)

Funding: This material is based upon work supported by the National Science Foundation (1915544) and the National Institute of General Medical Sciences of the National Institutes of Health (R25GM130517) to the Genomics Education Partnership (GEP; <https://thegep.org/>; 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. Supported by National Institutes of Health (United States) R25GM130517 to LK Reed. ,Supported by National Science Foundation (United States) 1915544 to LK Reed.

Author Contributions: Madeline L. Gruys: formal analysis, validation. Madison A. Sharp: formal analysis, validation, writing - original draft, writing - review editing. Zachary Lill: formal analysis, writing - review editing. Caroline Xiong: formal analysis, writing - review editing. Amy T. Hark: supervision, writing - review editing. James J. Youngblom: supervision, writing - review editing. Chinmay P. Rele: data curation, formal analysis, methodology, project administration, software, supervision, validation, visualization, writing - review editing. Laura K Reed: supervision, funding acquisition, conceptualization, project administration, writing - review editing.

Reviewed By: Anonymous

Nomenclature Validated By: Anonymous

History: Received December 19, 2023 **Revision Received** January 2, 2025 **Accepted** January 2, 2025 **Published Online** January 7, 2025 **Indexed** January 21, 2025

Copyright: © 2025 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: Gruys, ML; Sharp, MA; Lill, Z; Xiong, C; Hark, AT; Youngblom, JJ; Rele, CP; Reed, LK (2025). Gene model for the ortholog of *Glys* in *Drosophila simulans*. *microPublication Biology*. [10.17912/micropub.biology.001168](https://doi.org/10.17912/micropub.biology.001168)