

## Gene model for the ortholog of *gig* in *Drosophila ananassae*

Anne E. Backlund<sup>1</sup>, Logan Cohen<sup>2</sup>, Gregory Sileo<sup>3</sup>, Ian Alberts<sup>4</sup>, Thomas C. Giarla<sup>3</sup>, Chinmay P. Rele<sup>1</sup>, Laura K. Reed<sup>1§</sup>

<sup>1</sup>The University of Alabama, Tuscaloosa, AL USA

<sup>2</sup>Worcester State University, Worcester MA, USA

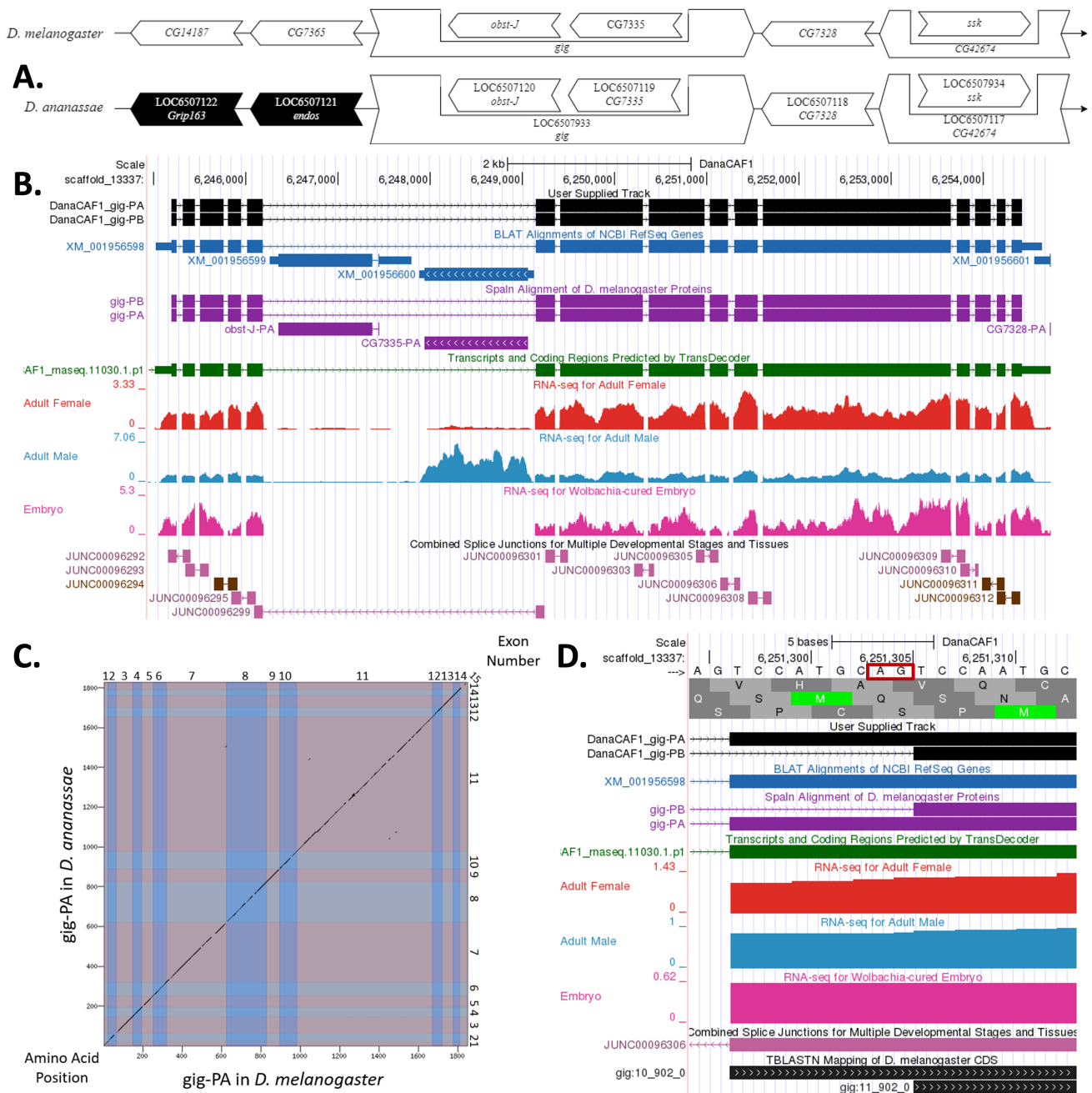
<sup>3</sup>Siena College, Loudonville, NY USA

<sup>4</sup>University of Evansville, Evansville IN USA

<sup>§</sup>To whom correspondence should be addressed: lreed1@ua.edu

### Abstract

Gene model for the ortholog of *gigas* (*gig*) in the *D. ananassae* May 2011 (Agencourt dana\_caf1/DanaCAF1) Genome Assembly (GenBank Accession: [GCA\\_000005115.1](#)). 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. Genomic neighborhood and gene model for *gig* in *Drosophila ananassae*:**

**(A) Synteny comparison of the genomic neighborhoods for *gig* in *Drosophila melanogaster* and *D. ananassae*.** Thin underlying arrows indicate the DNA strand within which the target gene—*gig*—is located in *D. melanogaster* (top) and *D. ananassae* (bottom). The thin arrows pointing to the right indicate that *gig* is on the positive (+) strand in *D. ananassae* and *D. melanogaster*. The wide gene arrows pointing in the same direction as *gig* are on the same strand relative to the thin underlying arrows, while wide gene arrows pointing in the opposite direction of *gig* are on the opposite strand relative to the thin underlying arrows. White gene arrows in *D. ananassae* indicate orthology to the corresponding gene in *D. melanogaster*, while black gene arrows indicate non-orthology. Gene symbols given in the *D. ananassae* gene arrows indicate the orthologous gene in *D. melanogaster*, while the locus identifiers are specific to *D. ananassae*. **(B) Gene Model in GEP UCSC Track Data Hub** (Raney et al., 2014). The coding-regions of *gig* in *D. ananassae* are displayed in the User Supplied Track (black); coding 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. ananassae*), 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, Adult Males, and *Wolbachia* cured Embryos (red, light blue, and pink, respectively; alignment of Illumina RNA-Seq reads from *D. ananassae*), and Splice Junctions Predicted by regtools using *D. ananassae* RNA-Seq ([SRP006203](#), [SRP007906](#), [PRJNA257286](#), [PRJNA388952](#)). Splice junctions shown have read-depths of 100-499 and 500-999 supporting reads in pink or brown, respectively. (C) **Dot Plot of gig-PA in *D. melanogaster* (x-axis) vs. the orthologous peptide in *D. ananassae* (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 mismatching amino acids at the specified location between species. (D) **Tenth CDS of Gene Model in GEP UCSC Track Data Hub** (Raney et al., 2014). The same evidence tracks as Figure 1B are displayed, in addition to the TBLASTN Mapping of *D. melanogaster* CDS track. Isoforms *gig-RA* and *gig-RB* differ only by the length of the tenth coding CDS: *gig-RB* is slightly shorter than *gig-RA*. Although there is only one prediction in the BLAT Alignment of NCBI RefSeq Genes track, the presence of the splice acceptor (AG) in the conserved position (red box) suggests that isoform *gig-RB* also exists in *D. ananassae*.

## 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](http://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).

"Gene *gig* (short for gigas, aka *TSC2*, *CG6975*, *FBgn0005198*) encodes a tumor suppressor protein that controls cell size, cell proliferation, and organ size (Ito and Rubin 1999). The *gig* protein contains a GTPase-activating protein (GAP) domain and forms a complex with protein *Tsc1* (Gao et al., 2001; Potter et al., 2001). The *gig-Tsc1* protein complex promotes GTP hydrolysis of the small G-protein *Rheb* (Ras homolog enriched in brain), thereby antagonizing the insulin and TOR signaling pathways (Gao et al., 2002; Zhang et al., 2003). This gene was originally identified in humans as *TSC2*, and mutations in *TSC1* or *TSC2* result in tuberous sclerosis, a syndrome characterized by widespread benign tumors (European Consortium 1993; van Sleightenhorst et al., 1997)." (Gruys et al., 2025b).

"*D. ananassae* (NCBI:txid7217) 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 Doleschall (1858). *D. ananassae* is circumtropical (Markow and O'Grady 2005; <https://www.taxodros.uzh.ch>, accessed 1 Feb 2023), and often associated with human settlement (Singh 2010). It has been extensively studied as a model for its cytogenetic and genetic characteristics, and in experimental evolution (Kikkawa 1938; Singh and Yadav 2015)." (Lawson et al., 2024).

We propose a gene model for the *D. ananassae* ortholog of the *D. melanogaster* *gigas* (*gig*) gene. The genomic region of the ortholog corresponds to the uncharacterized protein [XP\\_001956634.1](#) (Locus ID [LOC6507933](#)) in the May 2011 (Agencourt dana\_caf1/DanaCAF1) Genome Assembly of *D. ananassae* ([GCA\\_000005115.1](#); *Drosophila* 12 Genomes Consortium et al., 2007). This model is based on RNA-Seq data from *D. ananassae* ([SRP006203](#), [SRP007906](#), [PRJNA257286](#), [PRJNA388952](#); Graveley et al., 2011) and *gig* in *D. melanogaster* using FlyBase release *FB2023\_03* ([GCA\\_000001215.4](#); Larkin et al., 2021; Gramates et al., 2022).

### Syteny

The reference gene, *gig*, occurs on chromosome 3L in *D. melanogaster* and is flanked upstream by [CG14187](#) and [CG7365](#) and downstream by [CG7328](#) and [CG42674](#). The genes *obstructor-J* (*obst-J*) and [CG7335](#) are nested within *gig*, and *Snakeskin* (*Ssk*) is nested within [CG42674](#). The *tblastn* search of *D. melanogaster* *gig*-PA (query) against the *D. ananassae* (GCA000005115.1) Genome Assembly (database) placed the putative ortholog of *gig* within scaffold\_13337 at locus [LOC6507933](#) ([XP\\_001956634.1](#))— with an E-value of 0.0 and a percent identity of 68.20%. Furthermore, the putative ortholog is flanked upstream by [LOC6507122](#) ([XP\\_001956632.1](#)) and [LOC6507121](#) ([XP\\_044570979.1](#)), which correspond to *Grip163* and *endosulfine* (*endos*) in *D. melanogaster* (E-value: 0.0 and 3e-83; identity: 54.05% and 97.48%, respectively, as determined by *blastp*; Figure 1A; Altschul et al., 1990). The putative ortholog of *gig* is flanked downstream by [LOC6507118](#) ([XP\\_001956637.1](#)) and [LOC6507117](#) ([XP\\_032310418.1](#)), which correspond to [CG7328](#) and [CG42674](#) in *D. melanogaster* (E-value: 1e-161 and 0.0; identity: 68.54% and 82.06%, respectively, as determined by *blastp*). [LOC6507120](#) ([XP\\_001956635.1](#)) and [LOC6507119](#) ([XP\\_001956636.1](#)) are nested within *gig* (E-value: 1e-143 and 2e-170; identity: 53.95% and 65.33%). [LOC6507934](#) ([XP\\_001956640.1](#)) is nested within [CG42674](#) (E-value: 8e-115, identity: 98.77%). The putative ortholog assignment for *gig* in *D. ananassae* is supported by the following evidence: The genes downstream of and nested within the *gig* ortholog are orthologous to the genes at the same locus in *D. melanogaster* and local syteny is mostly conserved, supported by results generated from *blastp*, so we conclude that [LOC6507933](#) is the correct ortholog of *gig* in *D. ananassae* (Figure 1A).

### Protein Model

*gig* in *D. melanogaster* has two protein-coding isoforms (*gig*-PA and *gig*-PB; Figure 1B). Both mRNA isoforms contain fifteen protein-coding CDSs, but *gig*-RB has a slightly shorter tenth CDS than *gig*-RA. The isoform and protein-coding CDS counts are conserved in *D. ananassae*. The sequence of *gig*-PA in *D. ananassae* has 87.67% identity (E-value: 0.0) with the protein-coding isoform *gig*-PA in *D. melanogaster*, as determined by *blastp* (Figure 1C). Coordinates of this curated gene model (*gig*-PB, *gig*-PA) are stored by NCBI at GenBank/BankIt (accession [BK064615](#), [BK064616](#), respectively). These data are also archived in the CaltechDATA repository (see “Extended Data” section below).

### Special characteristics of the protein model

Although there is only one prediction in the BLAT Alignment of NCBI RefSeq Genes track (Figure 1B), there is evidence that both isoforms of *gig* are present in *D. ananassae*. *gig*-RA and *gig*-RB differ only by the length of their tenth CDS: *gig*-RB's CDS is slightly shorter than *gig*-RA. There is a splice acceptor site (AG) in the same reading frame that *gig*-RA uses (Figure 1D), which suggests that isoform *gig*-RB exists in *D. ananassae*. The TBLASTN Mapping of *D. melanogaster* CDS track also supports the presence of *gig*-RB.

### 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 syteny 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/genechecker/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 (Gruys et al., 2025a).

**Acknowledgements:** We would like to thank Wilson Leung for developing and maintaining the technological infrastructure that was used to create this gene model. Thank you to FlyBase for providing the definitive database for *Drosophila melanogaster* gene models. Also, thank you to Madeline Gruys for assistance in updating the manuscript to the current template.

## Extended Data

Description: Zip file containing FASTA, PEP, and GFF of the model. Resource Type: Model. File: [DanaCAF1\\_gig.zip](#). DOI: [10.22002/am768-qep80](https://doi.org/10.22002/am768-qep80)

## 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](#)
- Bock IR, Wheeler MR. (1972). The *Drosophila melanogaster* species group. *Univ. Texas Publs Stud. Genet.* 7(7213): 1-102. FBrf0024428
- Doleschall CL. 1858. Derde bijdrage tot de kennis der Dipteren fauna van nederlandsch indie. *Natuurk. Tijds. Ned.-Indie* 17: 73-128. FBrf0000091
- 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](#)
- European Chromosome 16 Tuberous Sclerosis Consortium. 1993. Identification and characterization of the tuberous sclerosis gene on chromosome 16. *Cell* 75(7): 1305-15. PubMed ID: [8269512](#)
- Gao X, Pan D. 2001. TSC1 and TSC2 tumor suppressors antagonize insulin signaling in cell growth. *Genes Dev* 15(11): 1383-92. PubMed ID: [11390358](#)
- Gao X, Zhang Y, Arrazola P, Hino O, Kobayashi T, Yeung RS, Ru B, Pan D. 2002. Tsc tumour suppressor proteins antagonize amino-acid-TOR signalling. *Nat Cell Biol* 4(9): 699-704. PubMed ID: [12172555](#)
- Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, dos Santos G, et al., Lovato. 2022. FlyBase: a guided tour of highlighted features. *Genetics* 220: 10.1093/genetics/iyac035. DOI: [10.1093/genetics/iyac035](https://doi.org/10.1093/genetics/iyac035)
- 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](#)
- Gruys ML, Sharp MA, Lill Z, Xiong C, Hark AT, Youngblom JJ, Rele CP, Reed LK. 2025a. Gene model for the ortholog of GlyS in *Drosophila simulans*. *MicroPubl Biol* 2025a: 10.17912/micropub.biology.001168. PubMed ID: [39845267](#)
- Gruys ML, Sharp MA, Croslyn C, Yap-Chiongco M, Reed LK, Stamm J, Rele CP. 2025b. Gene model for the ortholog of *gig* in *Drosophila mojavensis*. *microPublication Biology*. 10.17912/micropub.biology.001027. PubMed ID: [40838125](#)
- Hietakangas V, Cohen SM. 2009. Regulation of tissue growth through nutrient sensing. *Annu Rev Genet* 43: 389-410. PubMed ID: [19694515](#)
- Ito N, Rubin GM. 1999. *gigas*, a *Drosophila* homolog of tuberous sclerosis gene product-2, regulates the cell cycle. *Cell* 96(4): 529-39. PubMed ID: [10052455](#)
- 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](#)

Kikkawa H. 1938. Studies on the genetics and cytology of *Drosophila ananassae*. *Genetica* 20: 458-516. DOI: [10.1007/BF01531779](#)

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, McAbee M, Lucas RA, Tanner S, Wittke-Thompson J, Pelletier TA, et al., Reed LK. 2024. Gene model for the ortholog of *Ilp5* in *Drosophila ananassae*. *MicroPubl Biol* 2024. PubMed ID: [39717145](#)

Markow TA and O'Grady P. (2005) *Drosophila: A guide to species identification and use*. 978-0-12-473052-6

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](#)

Potter CJ, Huang H, Xu T. 2001. *Drosophila* Tsc1 functions with Tsc2 to antagonize insulin signaling in regulating cell growth, cell proliferation, and organ size. *Cell* 105(3): 357-68. PubMed ID: [11348592](#)

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. 2023. Manual annotation of *Drosophila* genes: a Genomics Education Partnership protocol. *F1000Research* 11: 1579. DOI: [10.12688/f1000research.126839.2](#)

Singh BN. 2010. *Drosophila ananassae*: a good model species for genetical, behavioural and evolutionary studies. *Indian J Exp Biol* 48(4): 333-45. PubMed ID: [20726331](#)

Singh BN, Yadav JP. 2015. Status of research on *Drosophila ananassae* at global level. *J Genet* 94(4): 785-92. PubMed ID: [26690536](#)

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](#)

van Slegtenhorst M, de Hoogt R, Hermans C, Nellist M, Janssen B, Verhoef S, et al., Kwiatkowski DJ. 1997. Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34. *Science* 277(5327): 805-8. PubMed ID: [9242607](#)

Zhang Y, Gao X, Saucedo LJ, Ru B, Edgar BA, Pan D. 2003. Rheb is a direct target of the tuberous sclerosis tumour suppressor proteins. *Nat Cell Biol* 5(6): 578-81. PubMed ID: [12771962](#)

**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). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health nor the National Science Foundation.

Supported by National Science Foundation (United States) 1915544 to LK Reed.

Supported by National Institutes of Health (United States) R25GM130517 to LK Reed.

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

**Author Contributions:** Anne E. Backlund: formal analysis, validation, writing - original draft, writing - review editing. Logan Cohen: data curation, validation, visualization, writing - review editing. Gregory Sileo: formal analysis, writing - review editing. Ian Alberts: formal analysis, writing - review editing. Thomas C. Giarla: supervision, writing - review editing. Chinmay P. Rele: data curation, formal analysis, methodology, project administration, software, supervision,

---

7/6/2026 - Open Access

validation, visualization, writing - review editing. Laura K. Reed: conceptualization, methodology, resources, validation, writing - review editing.

**Reviewed By:** Anonymous

**Nomenclature Validated By:** Anonymous

**History: Received** October 12, 2023 **Revision Received** June 26, 2026 **Accepted** July 3, 2026 **Published Online** July 6, 2026 **Indexed** July 20, 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:** Backlund AE, Cohen L, Sileo G, Alberts I, Giarla TC, Rele CP, Reed LK. 2026. Gene model for the ortholog of *gig* in *Drosophila ananassae*. microPublication Biology. [10.17912/micropub.biology.001031](https://doi.org/10.17912/micropub.biology.001031)