

Gene Model for the ortholog of *Ilp2* in *Drosophila ananassae*

Abigail R. Myers¹, Robyn Huber², Andrew M Arsham², Chinmay P. Rele¹, Laura Reed^{1§}

¹The University of Alabama, Tuscaloosa, AL USA

²Bemidji State University, Bemidji, MN USA

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

Abstract

Gene model for the ortholog of Insulin-like peptide 2 (*Ilp2*) in the *D. ananassae* May 2011 (Agencourt dana_caf1/DanaCAF1) Genome Assembly (GenBank Accession: [GCA_000005115.1](https://www.ncbi.nlm.nih.gov/nuccore/GCA_000005115.1)) of *Drosophila ananassae*. 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.

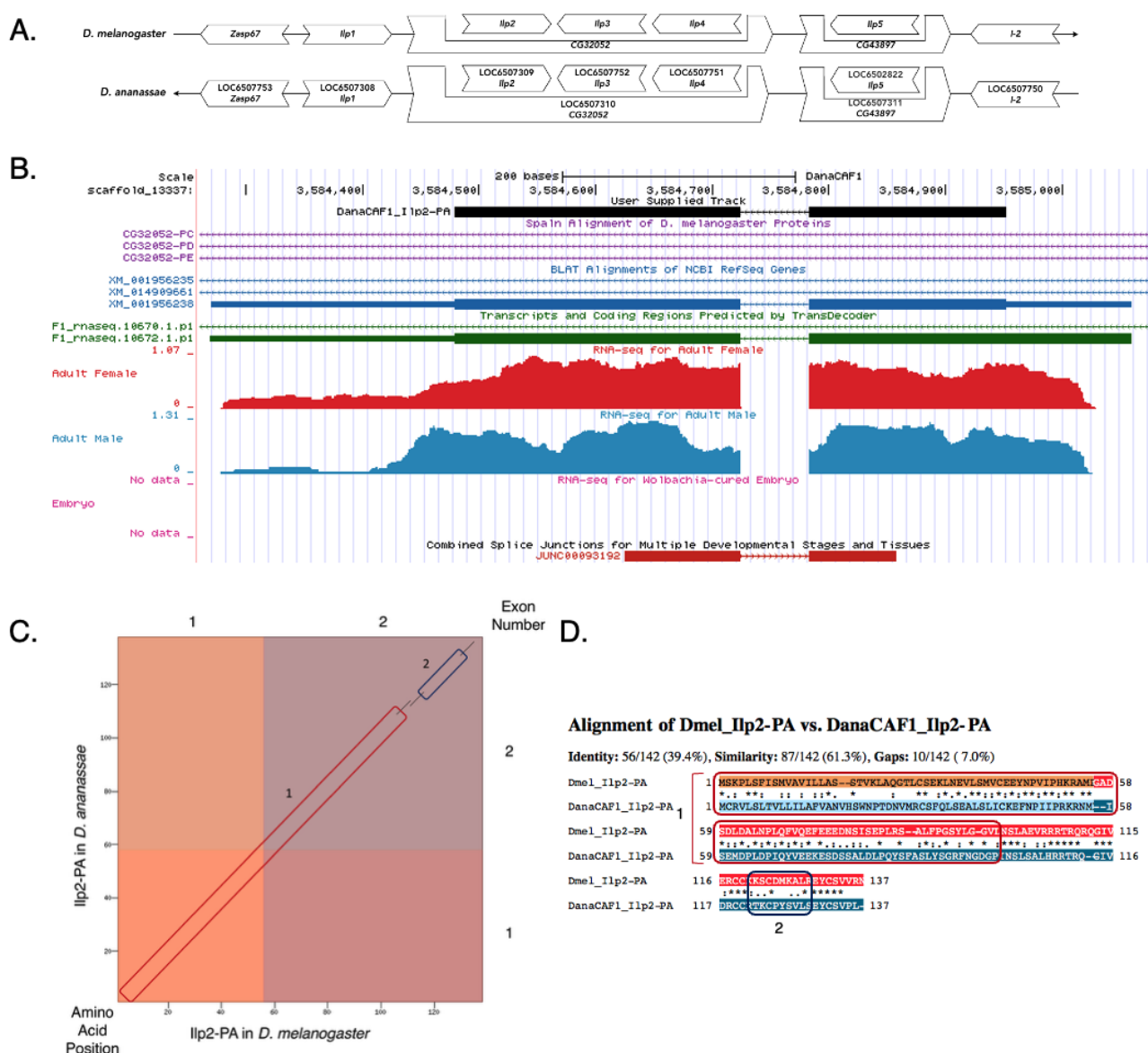


Figure 1. Genomic neighborhood and gene model for *Ilp2* in *D. ananassae*:

(A) A diagram of synteny of genomic neighborhood of *Ilp2* in *D. melanogaster* and *D. ananassae*. Gene arrows pointing in the same direction as *Ilp2* in both *D. ananassae* and *D. melanogaster* are on the same strand as *Ilp2*; while gene arrows pointing in the opposite direction are on the opposite strand. The thin underlying arrows pointing to the right indicate that

Ilp2 is on the + strand in *D. melanogaster*; arrows pointing to the left indicate that *Ilp2* is on the – strand in *D. ananassae*. White arrows in *D. ananassae* indicate the locus ID and the orthology to the corresponding gene in *D. melanogaster*. The gene names 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 UCSC Track Hub (Raney et al. 2014): the gene model in *D. ananassae* (black), Spaln of *D. melanogaster* Proteins (purple, alignment of refseq proteins from *D. melanogaster*), BLAT alignments of NCBI RefSeq Genes (blue, alignment of refseq genes for *D. ananassae*), RNA-Seq from Adult Females (red), Adult Males (blue), RNA-Seq for Wolbachia-cured Embryo (pink), alignment of Illumina RNAseq reads from *D. ananassae*, and Transcripts (green) including coding regions predicted by TransDecoder and Splice Junctions Predicted by regtools using *D. ananassae* RNA-Seq (Graveley et al., 2011; [SRP006203](#), [SRP007906](#), [PRJNA257286](#), [PRJNA388952](#)). The splice junction shown has a read-depth of 1777 with red supporting splice junctions having a range of >1000. The custom gene model (User Supplied Track) is indicated in black with CDS depicted with wide boxes, intron with narrow lines (arrows indicate direction of transcription). Note that the RNAseq for embryos shows no alignment of expression data at this locus, suggesting that this gene may not be expressed in embryos in this species, which neither supports or refutes the proposed model. Further note that the lack of an aligned Spaln *D. melanogaster* protein at this position indicates that the degree of sequence divergence between the reference gene and the target gene is greater than the minimum similarity needed to see an alignment for this algorithm. By default, Spaln is less sensitive and more specific than BLAST for assigning alignments. (C) Dot Plot of Ilp2-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; while CDS number is indicated along the top and right, and CDSs are also highlighted with alternating background colors. There are two large regions of sequence dissimilarity as displayed by the red (1) and blue (2) boxes. There is also one indel in the middle of CDS two represented by parallel lines. (D) The protein alignment of Ilp2-PA in *D. ananassae* against Ilp2-PA in *D. melanogaster* is shown. Boxes 1 and 2 correspond to the similarly labeled boxes in the Dot Plot highlighting regions of sequence dissimilarity.

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).

"Insulin-like peptide 2 (*Ilp2*), a core component of the insulin signaling pathway, mediates growth by acting as a ligand for the Insulin Receptor (*InR*) and transducing a signal via the Chico/PI3K/Akt(PKB) pathway (Brogiolo et al., 2001; Park et al., 2014). *Ilp2* plays a role in regulating body size by increasing the size and number of cells in individual organs (Ren et al., 2017). Evolutionary studies have shown that loss of *Ilp2* increases lifespan and changes in expression may have contributed to the evolution of body size in the Hawaiian *Drosophila* species (Grönke et al., 2010). In the absence of *Ilp2*, over-expression of *Ilp1* and *Ilp3-7* is enough to promote growth in *Drosophila* (Ikeya et al., 2002). *Ilp2* mutants also seem to have severe developmental delay (Grönke et al., 2010)." (Laskowski et al., 2022).

"*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).

The model presented here is the ortholog of *Ilp2* in the May 2011 (Agencourt dana_caf1/DanaCAF1) assembly of *D. ananassae* (Drosophila 12 Genomes Consortium 2007; [GCA_000005115.1](#)) and corresponds to the Gnomon Peptide ID ([XP_001956274.1](#)) predicted model in *D. ananassae* ([LOC6507309](#)). This gene model is based on RNA-Seq data from *D. ananassae* (Graveley et al, 2011; [SRP006203](#), [SRP007906](#), [PRJNA257286](#), [PRJNA388952](#)) and the *Ilp2* in *D. melanogaster* from FB2022_03 (Larkin et al., 2021; Gramates et al., 2022; Jenkins et al., 2022).

Synteny

Ilp2 occurs on Chromosome 3L in *D. melanogaster* and is flanked by *Zasp67* and *Ilp1* upstream. *Ilp2* is nested in [CG32052](#) along with *Ilp3* and *Ilp4* to the right. Downstream, *Ilp2* is flanked by [CG43897](#) (which nests *Ilp5*) and *I-2*. We determined that the putative ortholog of *Ilp2* is found on scaffold_13337 ([CH902618.1](#)) in *D. ananassae* with [LOC6507309](#) (via *tblastn* search with an e-value of 3e-18 and percent identity of 42.70%), where it is flanked by [LOC6507753](#) ([XP_014765450.1](#)) and [LOC6507308](#) ([XP_001956275.2](#)) which correspond to *Zasp67* and *Ilp1* in *D. melanogaster* with e-values 0.0 and 4e-24 and percent identities 67.67% and 52.46% respectively as determined by *blastp* (Figure 1A, Altschul et al., 1990). *Ilp2* is nested in [LOC6507310](#) ([XP_001956271.2](#)) which corresponds to [CG32052](#) in *D. melanogaster* with an e-value of 0.0 and a percent identity of 86.67% as determined by *blastp*. Nested in [CG32052](#) downstream of *Ilp2* are genes [LOC6507752](#) ([XP_001956273.1](#)) and [LOC6507751](#) ([XP_032309882.1](#)) which correspond to *Ilp3* and *Ilp4* in *D. melanogaster* with e-values of 1e-21 and 7e-28 and percent identities 46.32% and 48.91% respectively, as determined by *blastp*. Downstream of *Ilp2* is [LOC6507311](#) ([XP_044570593.1](#)) (which nests [LOC6502822](#) ([XP_001956270.1](#)) and [LOC6507750](#) ([XP_001956268.3](#)) which correspond to [CG43897](#), *Ilp5*, and *I-2* in *D. melanogaster* with e-values 0.0, 8e-09, and 1e-84 and percent identities 68.85%, 39.51%, and 72.50% respectively, as determined by *blastp*. We suggest this is the correct ortholog assignment for *Ilp2* in *D. ananassae* because local synteny is conserved and although there's a low percent similarity (46.79%) between *Ilp2*-PA in *D. ananassae* and *Ilp2*-PA in *D. melanogaster*.

Protein Model

Ilp2 in *D. ananassae* has one protein coding isoform (*Ilp2*-PA) (Figure 1B). Isoform (*Ilp2*-PA) contains two protein coding CDSs. Similarly, *Ilp2* in *D. melanogaster* has one protein coding isoform (*Ilp2*-PA) with two coding CDSs. The sequence of *Ilp2*-PA in *D. ananassae* has 46.79% identity with *Ilp2*-PA in *D. melanogaster* as determined by *blastp* (Figure 1C). There are large portions of sequence dissimilarity throughout the gene model as indicated by the red and blue boxes in the Dot Plot (Figure 1C). The coordinates of the curated gene models can be found in NCBI at GenBank/BankIt using the accession [BK064414](#). These data are also available in Extended Data files below, which are archived in CaltechData.

Special characteristics of the protein model

Sequence dissimilarity in gene model: The regions of sequence dissimilarity are highlighted with boxes in the Dot Plot (Figure 1C) and protein alignment (Figure 1D). Although these regions appear to be large, *Ilp2*-PA's overall short coding sequence amplifies the length of the gaps in the Dot Plot.

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/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., 2025).

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 and Logan Cohen for assistance in updating the manuscript to the current template.

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
- Brogiolo W, Stocker H, Ikeya T, Rintelen F, Fernandez R, Hafen E. 2001. An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr Biol* 11(4): 213-21. PubMed ID: [11250149](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- Grönke S, Clarke DF, Broughton S, Andrews TD, Partridge L. 2010. Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. *PLoS Genet* 6(2): e1000857. PubMed ID: [20195512](#)
- Gruys ML, Sharp MA, Lill Z, Xiong C, Hark AT, Youngblom JJ, Rele CP, Reed LK. 2025. Gene model for the ortholog of Glyc in *Drosophila simulans*. *MicroPubl Biol* 2025: 10.17912/micropub.biology.001168. PubMed ID: [39845267](#)
- Hietakangas V, Cohen SM. 2009. Regulation of tissue growth through nutrient sensing. *Annu Rev Genet* 43: 389-410. PubMed ID: [19694515](#)
- Hietakangas V, Cohen SM. 2009. Regulation of tissue growth through nutrient sensing. *Annu Rev Genet* 43: 389-410. PubMed ID: [19694515](#)
- Ikeya T, Galic M, Belawat P, Nairz K, Hafen E. 2002. Nutrient-dependent expression of insulin-like peptides from neuroendocrine cells in the CNS contributes to growth regulation in *Drosophila*. *Curr Biol* 12(15): 1293-300. PubMed ID: [12176357](#)
- 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](#)
- Laskowski, LF; Kiser, CA; Huber, R; Kusche, S; Arsham, AM; Giarla, TC; Rele, CP (2022). *Drosophila simulans* - *Ilp2*. *microPublication Biology*. DOI: [10.17912/micropub.biology.000679](#)
- 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

- 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](#)
- Mudge JM, Harrow J. 2016. The state of play in higher eukaryote gene annotation. Nat Rev Genet 17(12): 758-772. PubMed ID: [27773922](#)
- 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](#)
- Park S, Alfa RW, Topper SM, Kim GE, Kockel L, Kim SK. 2014. A genetic strategy to measure circulating *Drosophila* insulin reveals genes regulating insulin production and secretion. PLoS Genet 10(8): e1004555. PubMed ID: [25101872](#)
- 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](#)
- Ren S, Huang Z, Jiang Y, Wang T. 2018. dTBC1D7 regulates systemic growth independently of TSC through insulin signaling. J Cell Biol 217(2): 517-526. PubMed ID: [29187524](#)
- Singh BN, Yadav JP. 2015. Status of research on *Drosophila ananassae* at global level. J Genet 94(4): 785-92. PubMed ID: [26690536](#)
- 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](#)
- 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](#)

Funding: This material is based upon work supported by the National Science Foundation under Grant No. IUSE-1915544 to LKR and the National Institute of General Medical Sciences of the National Institutes of Health Award R25GM130517 to LKR. The Genomics Education Partnership is fully financed by Federal moneys. The content is solely the responsibility of the authors and does not necessarily represent the official views of 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.

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

Author Contributions: Abigail R. Myers: formal analysis, validation, writing - original draft, writing - review editing. Robyn Huber: formal analysis, writing - review editing. Andrew M Arsham: supervision, writing - review editing. Chinmay P. Rele: data curation, formal analysis, methodology, project administration, software, supervision, validation, visualization, writing - review editing. Laura Reed: conceptualization, funding acquisition, methodology, project administration, supervision, validation, visualization.

Reviewed By: Anonymous, David Molik

Nomenclature Validated By: Anonymous

History: Received July 7, 2023 **Revision Received** August 20, 2026 **Accepted** August 27, 2026 **Published Online** August 29, 2026 **Indexed** September 12, 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: Myers AR, Huber R, Arsham AM, Rele CP, Reed L. 2026. Gene Model for the ortholog of *Ilp2* in *Drosophila ananassae*. microPublication Biology. [10.17912/micropub.biology.000917](#)