

Gene model for the ortholog of *wrd* in *Drosophila ananassae*

Megan E. Lawson¹, Samantha Hoffman¹, Mikhail Sanu², Daniel Morris², Evan Merkhofer³, Stephanie Toering Peters⁴, Nikolaos Tsoதாக², Chinmay P. Rele¹, Laura K. Reed^{1§}

¹The University of Alabama, Tuscaloosa, AL USA

²Penn State Harrisburg, Middletown, PA USA

³Mount Saint Mary College, Newburgh, NY USA

⁴Wartburg College, Waverly, IA USA

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

Abstract

Gene model for the ortholog of *well-rounded* (*wrd*) in the 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.

4/1/2025 - Open Access

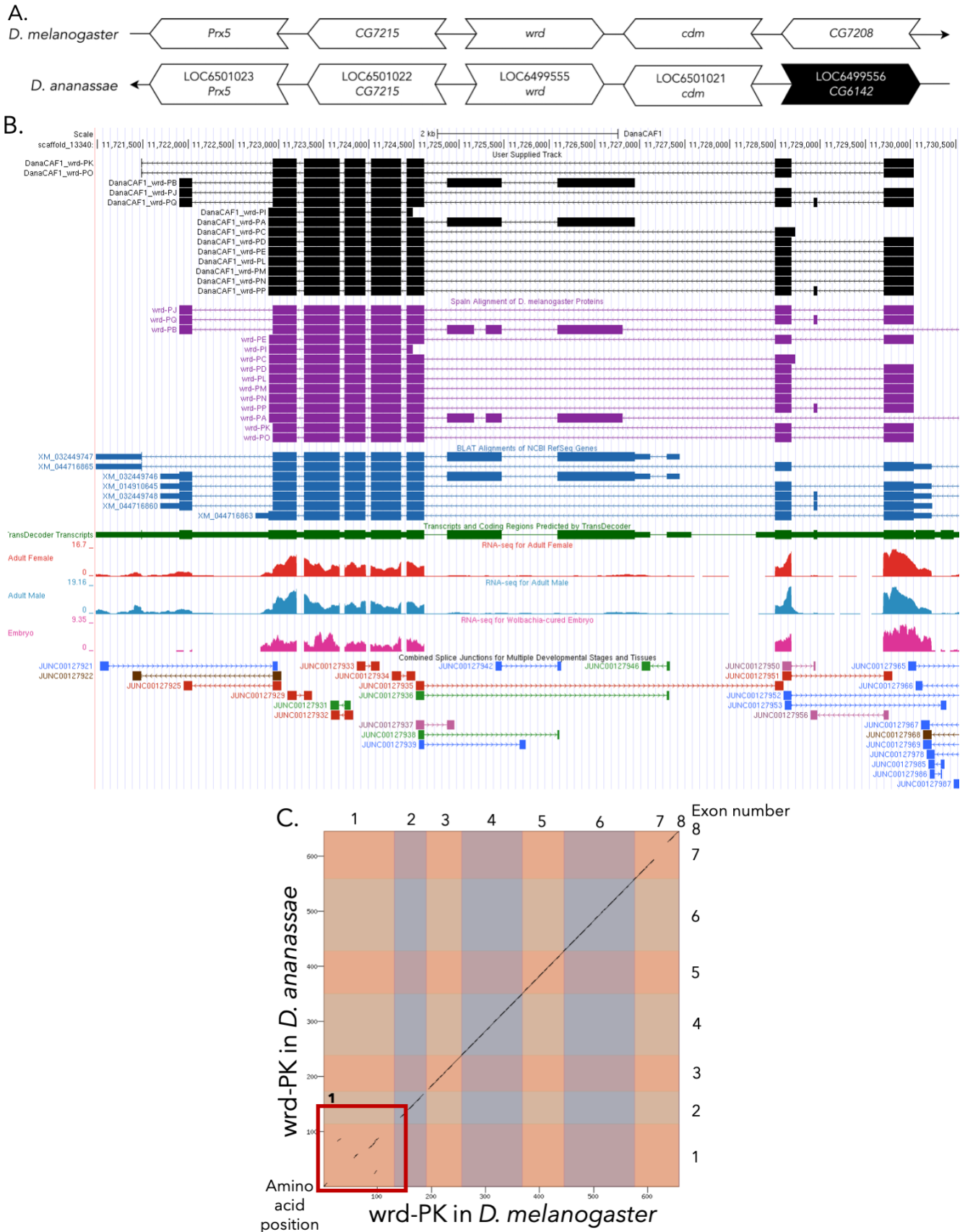


Figure 1. *wrd* gene model comparison between *Drosophila ananassae* and *Drosophila melanogaster* orthologs:

(A) Synteny of genomic neighborhood of *wrd* in *D. melanogaster* and *D. ananassae*. Gene arrows pointing in the same direction as reference gene in both *D. ananassae* and *D. melanogaster* are on the same strand as the reference gene; gene arrows pointing in the opposite direction are on the opposite strand. The thin underlying arrows pointing to the right indicate that *wrd* is on the + strand; arrows pointing to the left indicate that *wrd* is on the – strands. White arrows in *D. ananassae* indicate the locus ID and the orthology to the corresponding gene in *D. melanogaster* and black arrows indicate a non-orthologous gene. 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), and embryos (pink) (alignment of Illumina RNA-Seq reads from *D. ananassae*), and Transcripts (green) including coding regions predicted by TransDecoder and Splice Junctions Predicted by regtools using *D. ananassae* RNA-Seq (Gravely *et al.*, 2011; [SRP006203](#), [PRJNA257286](#), [SRP007906](#), [PRJNA388952](#)). Splice junctions shown have a minimum read-depth of 16 with 10-49, 50-99, 100-499, 500-999, >1000 supporting reads in blue, green, pink, brown, and red respectively. The custom gene model (User Supplied Track) is indicated in black with CDSs depicted with boxes and intron with narrow lines (arrows indicate direction of transcription). (C) Dot Plot of *wrd*-PK 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. Box 1 highlights a region containing some repeats and low sequence similarity.

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. 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 Doeschall (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 *wrd* in the May 2011 (Agencourt dana_caf1/DanaCAF1) assembly of *D. ananassae* ([GCA_000005115.1](#) – *Drosophila* 12 Genomes Consortium *et al.*, 2007) and corresponds to the Gnomon Peptide ID ([XP_032305638.1](#)) predicted model in *D. ananassae* ([LOC6499555](#)). This gene model is based on RNA-Seq data from *D. ananassae* (Gravely *et al.*, 2011; [SRP006203](#), [PRJNA257286](#), [SRP007906](#), [PRJNA388952](#)) and the *wrd* ([GCA_000001215.4](#)) in *D. melanogaster* from FB2022_03 (Larkin *et al.*, 2021; Gramates *et al.*, 2022; Jenkins *et al.*, 2022).

The *well-rounded* (*wrd*) gene was first identified in a gain-of-function screen for molecules that regulate synaptic development in the neuromuscular junction (NMJ) in *Drosophila melanogaster* (Viquez *et al.*, 2006). Its gene product is one of the two B' regulatory subunits of protein phosphatase 2A (PP2A), which interacts with liprin- α to stabilize specification of the active zone, a site in the axonal plasma membrane that regulates endo- and exocytosis of synaptic vesicles (Li *et al.*, 2014). Another

interacting partner of wrd is S6K, an important kinase in the insulin/TOR signaling pathway that is targeted for dephosphorylation by the PP2A holoenzyme through its molecular interaction with wrd (Hahn et al., 2010). Loss-of-function wrd mutants are viable and have fewer synaptic boutons that are larger and have a smooth round contour, a phenotype which gave the gene its name (Viquez et al., 2006). Knockout flies for the wrd B' subunit of PP2A are lean, have a reduced lifespan, and elevated insulin signaling (Hahn et al., 2010), with tissue specificity fine-tuned by cyclin G (Fischer et al., 2016).

Synteny

wrd occurs on chromosome 3L in *D. melanogaster* and is flanked upstream by [CG7215](#) and *Peroxiredoxin 5* (*Prx5*), and downstream by *cadmus* (*cdm*) and [CG7208](#). We determined that the putative ortholog of wrd is found on scaffold 13340 ([CH902617.1](#)) in *D. ananassae* with [LOC6499555](#) ([XP_032305638.1](#)) (via *tblastn* search with an e-value of 0.0 and percent identity of 88.81%). The wrd ortholog is flanked upstream by [LOC6501022](#) ([XP_001954393.1](#)) and [LOC6501023](#) ([XP_001954394.2](#)) which correspond to [CG7215](#) and *Prx5* in *D. melanogaster* with e-values of 7e-78 and 3e-127, respectively, and percent identities 82.31% and 93.68%, respectively, as determined by *blastp*. The wrd ortholog is flanked downstream by [LOC6501021](#) ([XP_001954390.2](#)) and [LOC6499556](#) ([XP_001954389.1](#)) which correspond to *cdm* and [CG6142](#) in *D. melanogaster* with e-values of 0.0 and 0.0, respectively, and percent identities of 87.44% and 88.64%, respectively, as determined by *blastp* (Figure 1A, Altschul et al., 1990). We believe this is the correct ortholog assignment for wrd in *D. ananassae* because synteny is well-conserved, with only one gene in the genomic neighborhood being non-orthologous, and because all of the BLAST searches used to determine orthology were very high quality.

Protein Model

wrd in *D. ananassae* has nine unique protein-coding isoforms (Figure 1B). The protein isoform encoded by mRNAs *wrd-RD*, *wrd-RE*, *wrd-RM*, *wrd-RL*, and *wrd-RN* (which differ in their UTRs) contains seven CDSs. The protein isoform encoded by mRNAs *wrd-RK* and *wrd-RO* contain eight protein-coding CDSs. Isoform wrd-PJ, is encoded by *wrd-RJ* which contains eight CDSs. Protein isoform wrd-PI is encoded by five CDSs. Isoform wrd-PQ is encoded by nine CDSs. Isoform wrd-PP is encoded by eight CDSs. Isoform wrd-PC is encoded by six CDSs. Isoform wrd-PB is encoded by eight CDSs. Isoform wrd-PA is encoded by seven CDSs. This isoform structure is the same relative to the ortholog in *D. melanogaster*, which also has 9 unique protein-coding isoforms. All of the *D. melanogaster* isoforms have the same number of CDSs as their corresponding ortholog in *D. ananassae*. The amino acid sequence of wrd in *D. ananassae* has 83.0% identity with the wrd in *D. melanogaster* as determined by *blastp* (Figure 1C). The coordinates of the curated gene models can be found in NCBI at GenBank/BankIt using the accessions [BK064428](#), [BK064429](#), [BK064430](#), [BK064431](#), [BK064432](#), [BK064433](#), [BK064434](#), [BK064435](#), [BK064436](#), [BK064437](#), [BK064438](#), [BK064439](#), [BK064440](#), and [BK064441](#). These data are also available in Extended Data files below, which are archived in CaltechData.

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

Extended Data

Description: GFF, FASTA, and PEP of the model. Resource Type: Model. File: [DanaCAF1_wrd.zip](#). DOI: [10.22002/9mj0d-yn206](#)

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990. Basic local alignment search tool. *J Mol Biol.* 215: 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 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 450: 03-18. PubMed ID: [17994087](#)
- Fischer P, Preiss A, Nagel AC 2016. A triangular connection between Cyclin G, PP2A and Akt1 in the regulation of growth and metabolism in *Drosophila*. *Fly (Austin).* 10(1): 11-8. PubMed ID: [26980713](#)
- 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: iyac035. PubMed ID: [335266522](#)
- 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: 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 Biology.* vol. 41, no. 5: 1006–1010. PubMed ID: [18992839](#)
- Hahn K, Miranda M, Francis VA, Vendrell J, Zorzano A, Teleman AA 2010. PP2A regulatory subunit PP2A-B' counteracts S6K phosphorylation. *Cell Metab.* 11(5): 438-44. PubMed ID: [20444422](#)
- Hietakangas V, Cohen SM 2009. Regulation of Tissue Growth through Nutrient Sensing. *Genetics.* vol. 43, no. 1: 389–410. PubMed ID: [19694515](#)
- 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: 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, Goodman JL, Gramates LS, Millburn G, Strelets VB, Tabone CJ, Thurmond J, 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](#)
- Li L, Tian X, Zhu M, Bulgari D, Böhme MA, Goettfert F, et al., Wu C. 2014. *Drosophila* Syd-1, liprin- α , and protein phosphatase 2A B' subunit Wrd function in a linear pathway to prevent ectopic accumulation of synaptic materials in distal axons. *J Neurosci* 34(25): 8474-87. PubMed ID: [24948803](#)
- 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: 10.17912/micropub.biology.000856. 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](#)

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, 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](#)

Viquez NM, Li CR, Wairkar YP, DiAntonio A. 2006. The B' protein phosphatase 2A regulatory subunit well-rounded regulates synaptic growth and cytoskeletal stability at the *Drosophila* neuromuscular junction. *J Neurosci* 26(36): 9293-303. PubMed ID: [16957085](#)

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.

Author Contributions: Megan E. Lawson: formal analysis, validation, writing - original draft, writing - review editing. Samantha Hoffman: formal analysis, writing - review editing. Mikhail Sanu: formal analysis, writing - review editing. Daniel Morris: formal analysis, writing - review editing. Evan Merkhofer: supervision, writing - review editing. Stephanie Toering Peters: supervision, writing - review editing. Nikolaos Tsoதாகos: 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, writing - review editing, project administration, methodology, funding acquisition, conceptualization.

Reviewed By: Anonymous

Nomenclature Validated By: Anonymous

History: Received July 18, 2023 **Revision Received** March 30, 2025 **Accepted** March 13, 2025 **Published Online** April 1, 2025 **Indexed** April 15, 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: Lawson ME, Hoffman S, Sanu M, Morris D, Merkhofer E, Toering Peters S, et al., Reed LK. 2025. Gene model for the ortholog of *wrd* in *Drosophila ananassae*. *microPublication Biology*. [10.17912/micropub.biology.000928](#)