

Gene model for the ortholog of *dock* in *Drosophila eugracilis*

Gabriella N. Bicanovsky¹, Megan E. Lawson¹, Alexandria Groeneveld², Jordan Elkinton³, Eden Dowell³, S. F. Shelley-Tremblay¹, Stephanie Toering Peters⁴, Lindsey J. Long³, Anya Goodman⁵, Melinda A. Yang⁶, Chinmay P. Rele¹, Laura K. Reed^{1§}

¹The University of Alabama, Tuscaloosa, AL USA

²Wartburg College, Waverly, IA, USA

³Oklahoma Christian University, Edmond, OK USA

⁴Wartburg College, Waverly, IA USA

⁵California Polytechnic State University, San Luis Obispo, CA USA

⁶University of Richmond, Richmond, VA USA

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

Abstract

Gene model for the ortholog of dreadlocks (*dock*) in the *D. eugracilis* Apr. 2013 (BCM-HGSC/Deug_2.0) (DeugGB2) Genome Assembly (GenBank Accession: [GCA_000236325.2](#)) of *Drosophila eugracilis*. 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.

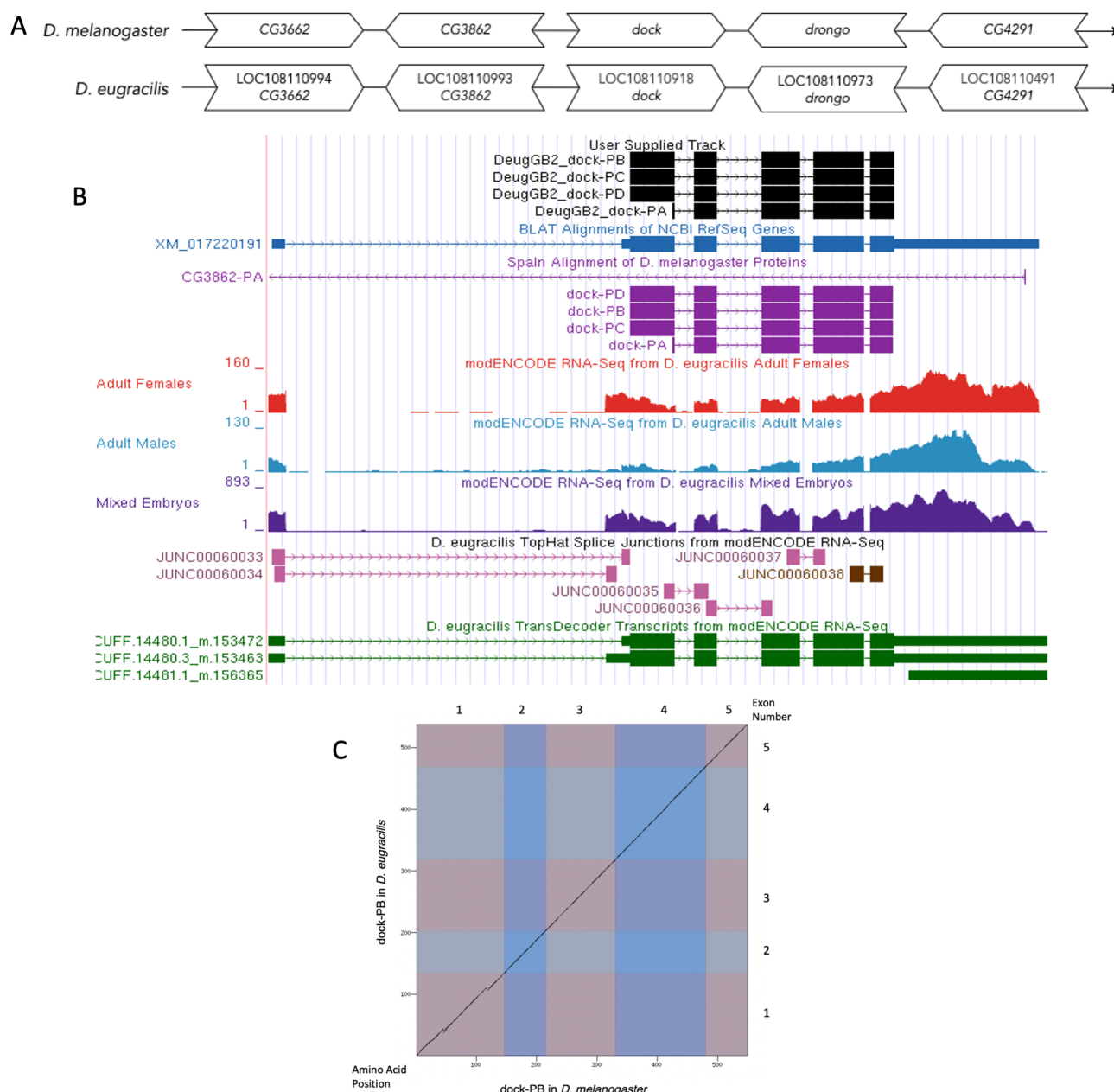


Figure 1. Genomic neighborhood and gene model for *dock* in *Drosophila eugracilis*:

(A) Synteny of genomic neighborhood of *dock* in *D. melanogaster* and *D. eugracilis*. Gene arrows pointing in the same direction as target gene in both *D. eugracilis* and *D. melanogaster* are on the same strand as the target gene; gene arrows pointing in the opposite direction are on the opposite strand. The thin underlying arrows pointing to the right indicate that *dock* is on the + strand. White arrows in *D. eugracilis* indicate the locus ID and the orthology to the corresponding gene in *D. melanogaster*. The gene names given in the *D. eugracilis* gene arrows indicate the orthologous gene in *D. melanogaster*, while the locus identifiers are specific to *D. eugracilis*. **(B) Gene Model in UCSC Track Hub** (Raney et al., 2014): the gene model in *D. eugracilis* (black), Spain 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. eugracilis*), RNA-Seq from adult females (red), adult males (blue), and mixed embryos (purple), alignment of Illumina RNA-Seq reads from *D. eugracilis*, and Transcripts (green) including coding regions predicted by TransDecoder and Splice Junctions Predicted by regtools using *D. eugracilis* RNA-Seq (PRJNA63469). Splice junctions shown have a minimum read-depth of 105 with 100-499 and 500-999 supporting reads in pink and brown, respectively. The custom gene model (User Supplied Track) is indicated in black with exon depicted with wide boxes, intron with narrow lines (arrows indicate direction of transcription). **(C) Dot Plot of dock-PB in *D. melanogaster* (x-axis) vs. the orthologous peptide in *D. eugracilis* (y-axis).** Amino acid number is indicated along the left and bottom; coding exon (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. eugracilis* (NCBI:txid29029) is part of the *melanogaster* species group within the subgenus *Sophophora* of the genus *Drosophila* (Pélandakis et al., 1993). It was first described as *Tanygastrella gracilis* by Duda (1924) and revised to *Drosophila eugracilis* by Bock and Wheeler (1972). *D. eugracilis* is found in humid tropical and subtropical forests across southeast Asia (<https://www.taxodros.uzh.ch>, accessed 1 Feb 2023)." (Morgan et al., 2022).

We propose a gene model for the *D. eugracilis* ortholog of the *D. melanogaster* *dreadlocks* (*dock*) gene. The genomic region of the ortholog corresponds to the uncharacterized protein [XP_017075680.1](#) (Locus ID [LOC108110918](#)) in the Apr. 2013 (BCM-HGSC.DEUG_2.0) Genome Assembly of *D. eugracilis* ([GCA_000236325.2](#)). This model is based on RNA-Seq data from *D. eugracilis* ([PRJNA63469](#); Chen et al., 2014) and *dock* in *D. melanogaster* using FlyBase release FB2023_03 ([GCA_000001215.4](#); Larkin et al., 2021; Gramates et al., 2022; Jenkins et al., 2022).

The product of gene *dreadlocks* (*dock*, FBgn0010583) is involved in several cellular functions, including axon guidance (Garritty et al., 1996; Hing et al., 1999; Schmucker et al., 2000; Stevens and Jacobs 2002; Weng et al., 2011), myoblast fusion during muscle fiber formation (Kaipa et al., 2013), regulation of intercellular bridges in germline cells during gametogenesis (Stark et al., 2021), and negative regulation of insulin receptor signaling pathway (Wu et al., 2011; Willoughby et al., 2017). A *dock* transcript was first isolated and sequenced in *Drosophila melanogaster* in a screen for P-element insertions leading to R cell projection defects (Garritty et al., 1996). *dock* encodes a protein that contains three N-terminal SH3 domains and one C-terminal SH2 domain known to bind to specific motifs and serve as binding adaptors (Garritty et al., 1996). Its regulation of photoreceptor axon guidance in *Drosophila* occurs through its interaction with InR (Rao et al., 1998; Song et al., 2003; Rao 2005). The dock protein plays a role in negatively regulating the insulin signaling pathway by facilitating the dephosphorylation of InR through recruitment of the ER-localized form of protein tyrosine phosphatase PTP61F, a function that is also observed in its mammalian ortholog *Nck* (Wu et al., 2011; Buszard et al., 2013).

Synteny

dock occurs on *chr2L* in *D. melanogaster* and is flanked by upstream genes [CG3662](#) and [CG3862](#) and downstream genes [drongo](#) (*drongo*) and [CG4291](#). It has been determined that the putative ortholog of *dock* is found on scaffold KB465133.1 in *D. eugracilis* (GB2 assembly [GCA_000236325.2](#)) with [LOC108110918](#) ([XP_017075680.1](#), via *tblastn* search with an e-value of 6e-164 and percent identity of 82.47%), where it is surrounded upstream by [LOC108110994](#) ([XP_017075791.1](#)) and [LOC108110993](#) ([XP_017075790.2](#)) which correspond to [CG3662](#) and [CG3862](#) in *D. melanogaster* with e-values of 0.0, and percent identity of 89.04% and 93.39%, respectively, as determined by *blastp* (Figure 1A, Altschul et al., 1990). The putative ortholog is flanked with downstream genes [LOC108110973](#) ([XP_017075764.2](#)) and [LOC108110491](#) ([XP_017075058.1](#)) which correspond to *drongo* and *CG4291* in *D. melanogaster* with e-values of 0.0, and percent identity of 85.61% and 86.98%, respectively, as determined by *blastp*. This is likely the correct ortholog assignment for *dock* in *D. eugracilis* for three reasons: first, a *tblastn* search with *D. melanogaster* dock-PA sequence against the *D. eugracilis* genome generated a top hit at this location with an E-value of 6e-164 and a percent identity of 82.475%. Second, a *blastp* search with the peptide sequence from *D. eugracilis* ([XP_017075680.1](#)) against the refseq_protein database in *D. melanogaster* generated a top hit of dock-PD with an e-value of 0.0 and percent identity of 87.43%. Third, the local synteny is highly conserved between the two genomic neighborhoods, with all the surrounding genes being orthologous to

each other. The high similarity between the genomic neighborhood and gene identity is expected as *D. eugracilis* is closely related to *D. melanogaster*.

Protein Model

dock in *D. eugracilis* has encodes two unique protein isoforms; one encoded by mRNAs *dock-RB*, *dock-RC*, and *dock-RD* that all have identical CDSs, while *dock-RA* is unique (Figure 1B). All mRNA isoforms contain five CDSs. Relative to the ortholog in *D. melanogaster*, the RNA CDS number and protein isoform count are conserved. When comparing the protein alignment of *dock-PB* between *D. eugracilis* and *D. melanogaster* there is a high level of protein similarity and an 87.43% amino acid identity as determined by *blastp* (Figure 1C). The coordinates of the curated gene models can be found in NCBI at GenBank/BankIt using the accessions [BK064454](#), [BK064455](#), [BK064456](#), and [BK064457](#). 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 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.

Extended Data

Description: A GFF, FASTA, and PEP of the model. Resource Type: Model. File: [DeugGB2_dock.zip](#). DOI: [10.22002/b9xm7-ren93](https://doi.org/10.22002/b9xm7-ren93)

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
- Buszard BJ, Johnson TK, Meng TC, Burke R, Warr CG, Tiganis T. 2013. The nucleus- and endoplasmic reticulum-targeted forms of protein tyrosine phosphatase 61F regulate *Drosophila* growth, life span, and fecundity. *Mol Cell Biol* 33(7): 1345-56. PubMed ID: [23339871](#)
- Chen ZX, Sturgill D, Qu J, Jiang H, Park S, Boley N, et al., Richards S. 2014. Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res* 24(7): 1209-23. PubMed ID: [24985915](#)

- Duda, O. 1924. Revision der europäischen u. grönländischen sowie einiger sudostasiat. Arten der Gattung Piophilala Fallén (Dipteren) [part]. Konowia 3: 97-113 [1924.07.10]
- Garritty PA, Rao Y, Salecker I, McGlade J, Pawson T, Zipursky SL. 1996. Drosophila photoreceptor axon guidance and targeting requires the dreadlocks SH2/SH3 adapter protein. Cell 85(5): 639-50. PubMed ID: [8646773](#)
- 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(4): 10.1093/genetics/iyac035. PubMed ID: [35266522](#)
- 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](#)
- 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](#)
- Hing H, Xiao J, Harden N, Lim L, Zipursky SL. 1999. Pak functions downstream of Dock to regulate photoreceptor axon guidance in Drosophila. Cell 97(7): 853-63. PubMed ID: [10399914](#)
- 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](#)
- Kaipa BR, Shao H, Schäfer G, Trinkewitz T, Groth V, Liu J, et al., Önel SF. 2013. Dock mediates Scar- and WASp-dependent actin polymerization through interaction with cell adhesion molecules in founder cells and fusion-competent myoblasts. J Cell Sci 126(Pt 1): 360-72. PubMed ID: [22992459](#)
- 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](#)
- 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](#)
- Morgan A, Kiser CA, Bronson I, Lin H, Guillelte N, McMahon R, et al., Rele CP. 2022. Drosophila eugracilis - Akt. MicroPubl Biol 2022: 10.17912/micropub.biology.000544. PubMed ID: [35856017](#)
- 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](#)
- Pélandakis M, Solignac M. 1993. Molecular phylogeny of Drosophila based on ribosomal RNA sequences. J Mol Evol 37(5): 525-43. PubMed ID: [8283482](#)
- 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](#)
- Rao Y, Zipursky SL. 1998. Domain requirements for the Dock adapter protein in growth- cone signaling. Proc Natl Acad Sci U S A 95(5): 2077-82. PubMed ID: [9482841](#)
- Rao Y. 2005. Dissecting Nck/Dock signaling pathways in Drosophila visual system. Int J Biol Sci 1(2): 80-6. PubMed ID: [15951852](#)
- 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](#)
- Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. 2000. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101(6): 671-84. PubMed ID: [10892653](#)
- Song J, Wu L, Chen Z, Kohanski RA, Pick L. 2003. Axons guided by insulin receptor in Drosophila visual system. Science 300(5618): 502-5. PubMed ID: [12702880](#)
- Stark K, Crowe O, Lewellyn L. 2021. Precise levels of the Drosophila adaptor protein Dreadlocks maintain the size and stability of germline ring canals. J Cell Sci 134(8). PubMed ID: [33912915](#)

Stevens A, Jacobs JR. 2002. Integrins regulate responsiveness to slit repellent signals. *J Neurosci* 22(11): 4448-55. PubMed ID: [12040052](#)

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](#)

Weng YL, Liu N, DiAntonio A, Broihier HT. 2011. The cytoplasmic adaptor protein Caskin mediates Lar signal transduction during *Drosophila* motor axon guidance. *J Neurosci* 31(12): 4421-33. PubMed ID: [21430143](#)

Willoughby LF, Manent J, Allan K, Lee H, Portela M, Wiede F, et al., Richardson HE. 2017. Differential regulation of protein tyrosine kinase signalling by Dock and the PTP61F variants. *FEBS J* 284(14): 2231-2250. PubMed ID: [28544778](#)

Wu CL, Buszard B, Teng CH, Chen WL, Warr CG, Tiganis T, Meng TC. 2011. Dock/Nck facilitates PTP61F/PTP1B regulation of insulin signalling. *Biochem J* 439(1): 151-9. PubMed ID: [21707536](#)

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.

Author Contributions: Gabriella N. Bicanovsky: formal analysis, validation, writing - original draft, writing - review editing. Megan E. Lawson: formal analysis, validation, writing - original draft, writing - review editing. Alexandria Groeneveld: formal analysis, writing - review editing. Jordan Elkinton: formal analysis, writing - review editing. Eden Dowell: formal analysis, writing - review editing. S. F. Shelley-Tremblay: formal analysis, writing - review editing. Stephanie Toering Peters: supervision, writing - review editing. Lindsey J. Long: supervision, writing - review editing. Anya Goodman: supervision, writing - review editing. Melinda A. Yang: 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, conceptualization, funding acquisition, methodology, project administration.

Reviewed By: Christopher Shaffer

Nomenclature Validated By: Anonymous

History: Received September 27, 2023 **Revision Received** September 3, 2025 **Accepted** September 5, 2025 **Published Online** September 11, 2025 **Indexed** September 25, 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: Bicanovsky GN, Lawson ME, Groeneveld A, Elkinton J, Dowell E, Shelley-Tremblay SF, et al., Reed LK. 2025. Gene model for the ortholog of *dock* in *Drosophila eugracilis*. *microPublication Biology*. [10.17912/micropub.biology.001848](#)