

Gene model for the ortholog of *Roc1a* in *Drosophila ananassae*

Megan E. Lawson¹, Kelsey Gammage², Calvin Dixel², Lindsey J. Long², Chinmay P. Rele¹, Laura K. Reed^{1§}

¹The University of Alabama, Tuscaloosa, AL USA

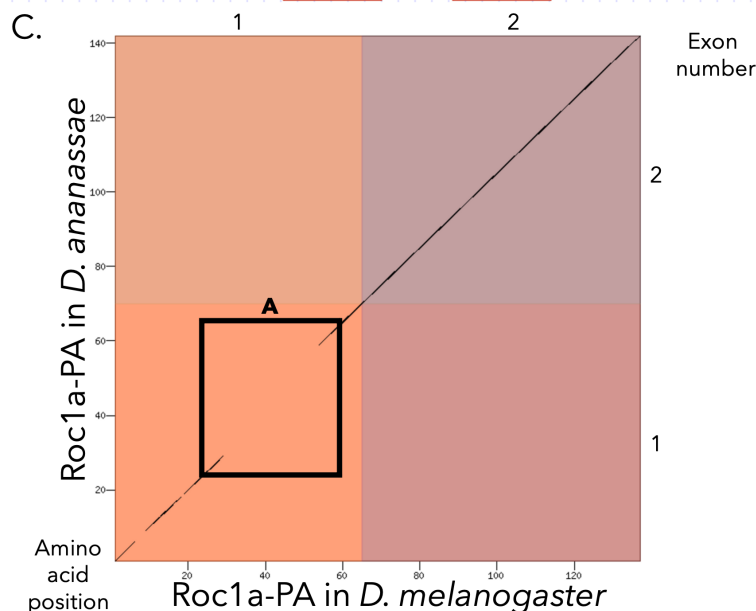
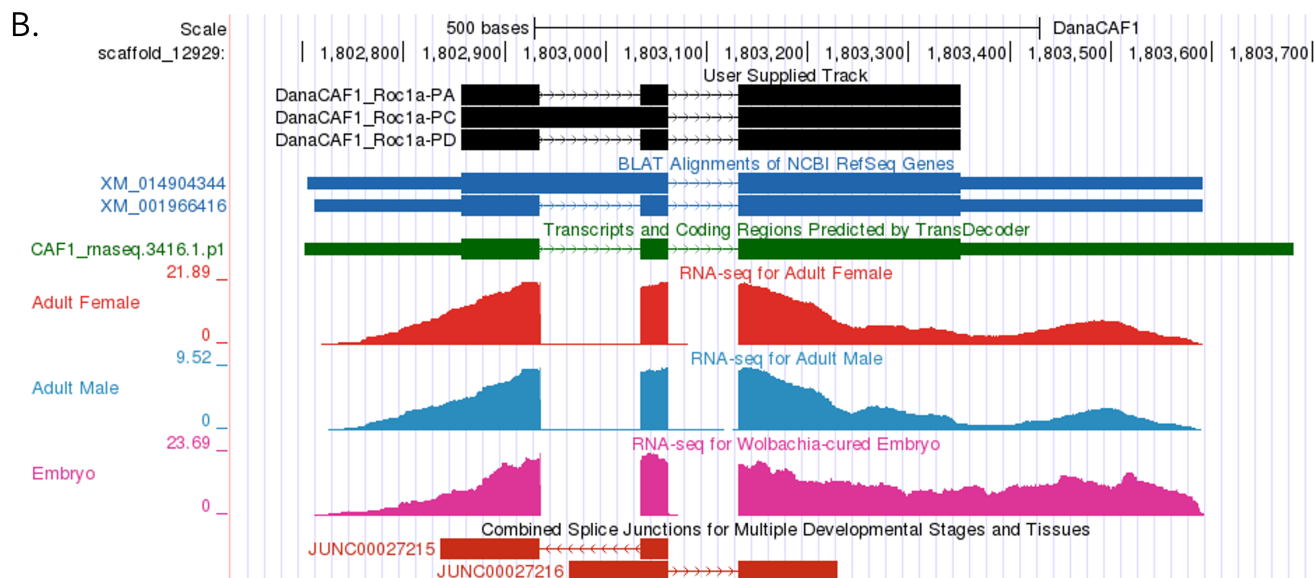
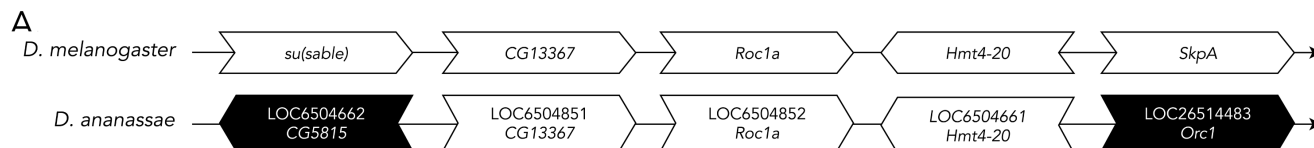
²Oklahoma Christian University, Edmond, OK USA

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

Abstract

Gene model for the ortholog of *Regulator of cullins 1a* ([Roc1a](#)) in the *Drosophila 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.

6/4/2025 - Open Access



D. Alignment of Dmel_Roc1a-PC vs. DanaCAF1_Roc1a-PC

Identity: 112/141 (79.4%), **Similarity:** 120/141 (85.1%), **Gaps:** 5/141 (3.5%)

B

Dmel_Roc1a-PC	1	MEVDEDGYEVPSSSSKGDKKRFEVKK	VS--GQOKS--RVIVNECTDGNSTSSFP-LRRKQW	55
DanaCAF1_Roc1a-PC	1	MEVDEEEFVPSSSSKGEKKRFEVKK	/SKLGEESAFLAVMAGPLLILQLSRFIFLGKMOW	60
Dmel_Roc1a-PC	56	NAVALWAWD	IVVDNCAICRNHIMDLCEIQANQASATSEECTVAVGVCNHAHFHFCISRW	115
DanaCAF1_Roc1a-PC	61	NAVALWAWD	IVVDNCAICRNHIMDLCEIQANQASATSEECTVAVGVCNHAHFHFCISRW	120
Dmel_Roc1a-PC	116	LKTRQVCPLDNREWDFQYGH		136
DanaCAF1_Roc1a-PC	121	LKTRQVCPLDNREWDFQYGH		141

Figure 1. *Roc1a* gene model comparison between *Drosophila ananassae* and *Drosophila melanogaster* orthologs:

(A) Synteny comparison of the genomic neighborhoods for *Roc1a* in *Drosophila melanogaster* and *D. ananassae*. Thin underlying arrows indicate the DNA strand within which the reference gene—*Roc1a*—is located in *D. melanogaster* (top) and *D. ananassae* (bottom). Thin arrows pointing to the right indicate that *Roc1a* is on the positive (+) strand in *D. melanogaster* and *D. ananassae*. The wide gene arrows pointing in the same direction as *Roc1a* are on the same strand relative to the thin underlying arrows, while wide gene arrows pointing in the opposite direction of *Roc1a* 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*. 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 *Roc1a* 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*), Transcripts and Coding Regions Predicted by TransDecoder (dark green), RNA-Seq from Adult Females, Adult Males, and Wolbachia- cured embryos (red, blue, and pink, respectively; alignment of Illumina RNA-Seq reads from *D. ananassae*), 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 10 with >1000 supporting reads indicated in red. **(C) Dot Plot of Roc1a-PC 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 A highlights a gap in the alignment caused by low sequence similarity between this region of Roc1a-PC in *D. ananassae* compared to *D. melanogaster*. **(D) Protein alignment of Roc1a-PC in *D. melanogaster* (top row) vs. the orthologous peptide in *D. ananassae* (bottom row).** The amino acids in Roc1a-RC are indicated in each species' respective row, with CDSs being highlighted in alternating colors. Asterisks between the protein alignment from each species indicate that the amino acid at that position is the same, colons and periods indicate the amino acids are very similar structurally (colons at a higher extent and periods at a lesser extent), and empty space indicates the amino acids are not similar. Lines in the protein sequence in either species indicate that there are gaps in that portion of the alignment for that species. Box 1D-B indicates the same region of decreased sequence similarity in CDS one of *Roc1a-RC* that is highlighted in Box 1C-A.

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).

"*Roc1a* (also known as *Rbx1*) is a member of the SCF E3 ubiquitin ligase complex and was originally identified through sequence similarity with vertebrate and yeast homologs and biochemical interaction studies (Bocca et al., 2001). *Roc1a* deletion mutants are lethal between the first and second larval instars and *Roc1a* mutant clones in imaginal discs have cell proliferation defects (Nouredine et al., 2002). In addition, *Roc1a* and other members of the SCF E3 ubiquitin ligase complex function in the pruning of larval neurons by targeting the insulin-responsive kinase Akt for ubiquitination and degradation, thus inhibiting insulin signaling (Wong et al., 2013)." (Lawson et al., 2025).

"*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).

We propose a gene model for the *D. ananassae* ortholog of the *D. melanogaster* *Regulator of cullins 1a* (*Roc1a*) gene. The genomic region of the ortholog corresponds to the uncharacterized protein [LOC6504852](#) (RefSeq accession [XP_014759830.1](#)) in the May 2011 (Agencourt dana_caf1/DanaCAF1) Genome Assembly of *D. ananassae* (GenBank Accession: [GCA_000005115.1](#); *Drosophila* 12 Genomes Consortium et al., 2007). This model is based on RNA-Seq data from *D. ananassae* (Gravely et al. 2011; [SRP006203](#), [PRJNA257286](#), [SRP007906](#), [PRJNA388952](#)) and *Roc1a* in *D. melanogaster* using FlyBase release FB2023_02 ([GCA_000001215.4](#); Larkin et al. 2021; Gramates et al., 2022; Jenkins et al., 2022).

Synteny

The reference gene, *Roc1a*, occurs on chromosome X in *D. melanogaster* and is flanked upstream by *CG13367* and *suppressor of sable* (*Su(sable)*) and downstream by *Histone methyltransferase 4-20* (*Hmt4-20*) and *SKP1-related A* (*SkpA*). The *tblastn* search of *D. melanogaster* *Roc1a*-PC (query) against the *D. ananassae* (GenBank Accession: [GCA_000005115.1](#)) Genome Assembly (database) placed the putative ortholog of *Roc1a* within scaffold_12929 ([CH902632.1](#)) at locus [LOC6504852](#) ([XP_014759830.1](#))— with an E-value of 2e-63 and a percent identity of 68.29%. Furthermore, the putative ortholog is flanked upstream by [LOC6504851](#) ([XP_001966451.1](#)) and [LOC6504662](#) ([XP_001966450.2](#)), which correspond to *CG13367* and *CG5815* in *D. melanogaster* (E-value: 8e-176 and 0.0; identity: 68.19% and 67.55%, respectively, as determined by *blastp*; Figure 1A, Altschul et al., 1990). The putative ortholog of *Roc1a* is flanked downstream by [LOC6504661](#) ([XP_032309073.1](#)) and [LOC26514483](#) ([XP_014759887.1](#)), which correspond to *Hmt4-20* and *Ocr1* in *D. melanogaster* (E-value: 0.0 and 0.0; identity: 67.40% and 69.68%, respectively, as determined by *blastp*). The putative ortholog assignment for *Roc1a* in *D. ananassae* is supported by the partial conservation of the synteny of this genomic neighborhood, and all BLAST results indicate good-quality matches.

Protein Model

Roc1a in *D. ananassae* has two unique protein-coding isoforms: *Roc1a*-PA (identical to *Roc1a*-PD) and *Roc1a*-PC (Figure 1B). mRNA isoforms *Roc1a*-RA and *Roc1a*-RD contain three protein-coding CDSs. mRNA isoform *Roc1a*-RC has two CDSs,

with its first CDS (FlyBase ID: 1_2094_0) spanning the length of the first two CDSs of *Roc1a-RA* and *Roc1a-RD* (FlyBase IDs: 2_2094_0 and 3_2094_0) combined (all CDS IDs based on FlyBase release FB2023_02; [GCA 000001215.4](https://flybase.org/reports/FB000012154)). Relative to the ortholog in *D. melanogaster*, the RNA CDS number and isoform structure is conserved. The sequence of *Roc1a-PC* in *D. ananassae* has 79.43% identity (E-value: 2e-76) with the protein-coding isoform *Roc1a-PC* in *D. melanogaster*, as determined by *blastp* (Figure 1C). Coordinates of this curated gene model (*Roc1a-PC*, *Roc1a-PA*, *Roc1a-PD*) are stored by NCBI at GenBank/BankIt (accession [BK064627](https://www.ncbi.nlm.nih.gov/nuccore/BK064627), [BK064628](https://www.ncbi.nlm.nih.gov/nuccore/BK064628), [BK064629](https://www.ncbi.nlm.nih.gov/nuccore/BK064629), respectively). These data are also archived in the CaltechDATA repository (see “Extended Data” section below).

Special characteristics of the protein model

Low sequence similarity in the latter portion of the first CDS of *Roc1a-RC*

Boxes 1C-A and 1D-B highlight a region of decreased sequence similarity in the alignment of the first CDS of *Roc1a-RC* in *D. melanogaster* and *D. ananassae*. As shown in the dot plot and protein alignment (Figure 1C-A; 1D-B), many of the dissimilarities between the *Roc1a-RC* protein sequence in *D. melanogaster* and *D. ananassae* are found in the latter portion of the first CDS, which is unique to only isoform *Roc1a-RC*, as isoforms *Roc1a-RA* and *Roc1a-RD* have spliced out this region. A *blastp* search of *Roc1a-PA* in *D. ananassae* has 96.30% identity (E-value: 5e-76) with the protein-coding isoform *Roc1a-PA* in *D. melanogaster*, which is much higher relative to the 79.43% identity of the same search using *Roc1a-PC*. The significantly higher identity value for the other isoforms of *Roc1a* in *D. ananassae*, taken in the context of the protein alignment shown in figure 1D, provides further evidence that this is likely the correct ortholog, and that while the portion of the protein alignment unique to *Roc1a-PC* contains low sequence similarity, the remaining regions of *Roc1a* are very well-conserved.

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: 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 Zip file containing a GFF, FASTA, and PEP of the model. Resource Type: Model. File: [DanaCAF1_Roc1a.zip](#). DOI: [10.22002/je5dq-jg936](https://doi.org/10.22002/je5dq-jg936)

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](#)
- Bocca SN, Muzzopappa M, Silberstein S, Wappner P. 2001. Occurrence of a putative SCF ubiquitin ligase complex in *Drosophila*. *Biochem Biophys Res Commun* 286(2): 357-64. PubMed ID: [11500045](#)
- Bock IR, Wheeler MR (1972). The *Drosophila melanogaster* species group. *Univ. Texas Publs Stud. Genet.* 7(7213): 1-102. FBrf0024428 PubMed ID: [null](#)
- Doleschall CL. (1858). Derde bijdrage tot de kennis der Dipteren fauna van nederlandsch indie. *Natuurk. Tijd. Ned.-Indie* 17: 73-128. FBrf0000091 PubMed ID: [null](#)
- 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. 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](#)
- 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. 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, Wellik IG, Alvarado B, German T, Thompson JS, Long LJ, et al., Reed LK. 2025. Gene model for the ortholog of *Roc1a* in *Drosophila eugracilis*. *MicroPubl Biol* 2025: 10.17912/micropub.biology.001028. PubMed ID: [40110471](#)
- 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, O'Grady P. (2005) *Drosophila: A guide to species identification and use*. 978-0-12-473052-6 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](#)
- Noureddine MA, Donaldson TD, Thacker SA, Duronio RJ. 2002. *Drosophila* *Roc1a* encodes a RING-H2 protein with a unique function in processing the Hh signal transducer Ci by the SCF E3 ubiquitin ligase. *Dev Cell* 2(6): 757-70. PubMed ID: [12062088](#)
- 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. PubMed ID: [null](#)

6/4/2025 - Open Access

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](#)

Wong JJ, Li S, Lim EK, Wang Y, Wang C, Zhang H, et al., Yu F. 2013. A Cullin1-based SCF E3 ubiquitin ligase targets the InR/PI3K/TOR pathway to regulate neuronal pruning. PLoS Biol 11(9): e1001657. PubMed ID: [24068890](#)

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; [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 Science Foundation (United States) 1915544 to LK Reed.

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

Author Contributions: Megan E. Lawson: formal analysis, validation, writing - original draft, writing - review editing. Kelsey Gammage: formal analysis, writing - review editing. Calvin Dixel: formal analysis, writing - review editing. Lindsey J. Long: 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, investigation, conceptualization, funding acquisition, methodology.

Reviewed By: Anonymous

History: Received October 21, 2023 **Revision Received** May 30, 2025 **Accepted** June 2, 2025 **Published Online** June 4, 2025 **Indexed** June 18, 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, Gammage K, Dixel C, Long LJ, Rele CP, Reed LK. 2025. Gene model for the ortholog of *Roc1a* in *Drosophila ananassae*. microPublication Biology. [10.17912/micropub.biology.001036](#)