microPublication BIOLOGY

2/13/2025 - *Open Access*

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

Megan E. Lawson[1], Ryan J. Dufur[2], Braden Wright[2], Isabel G. Wellik[3], Lindsey J. Long[2], Jeffrey S. Thompson[3], Chinmay P. Rele[1], Laura K Reed[1§]

[1]University of Alabama, Tuscaloosa, AL USA

[2]Oklahoma Christian University, Edmond, OK USA

[3]Denison University, Granville, OH 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 mojavensis* May 2011 (Agencourt dmoj_caf1/DmojCAF1) Genome Assembly (GenBank Accession: GCA_000005175.1). 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.

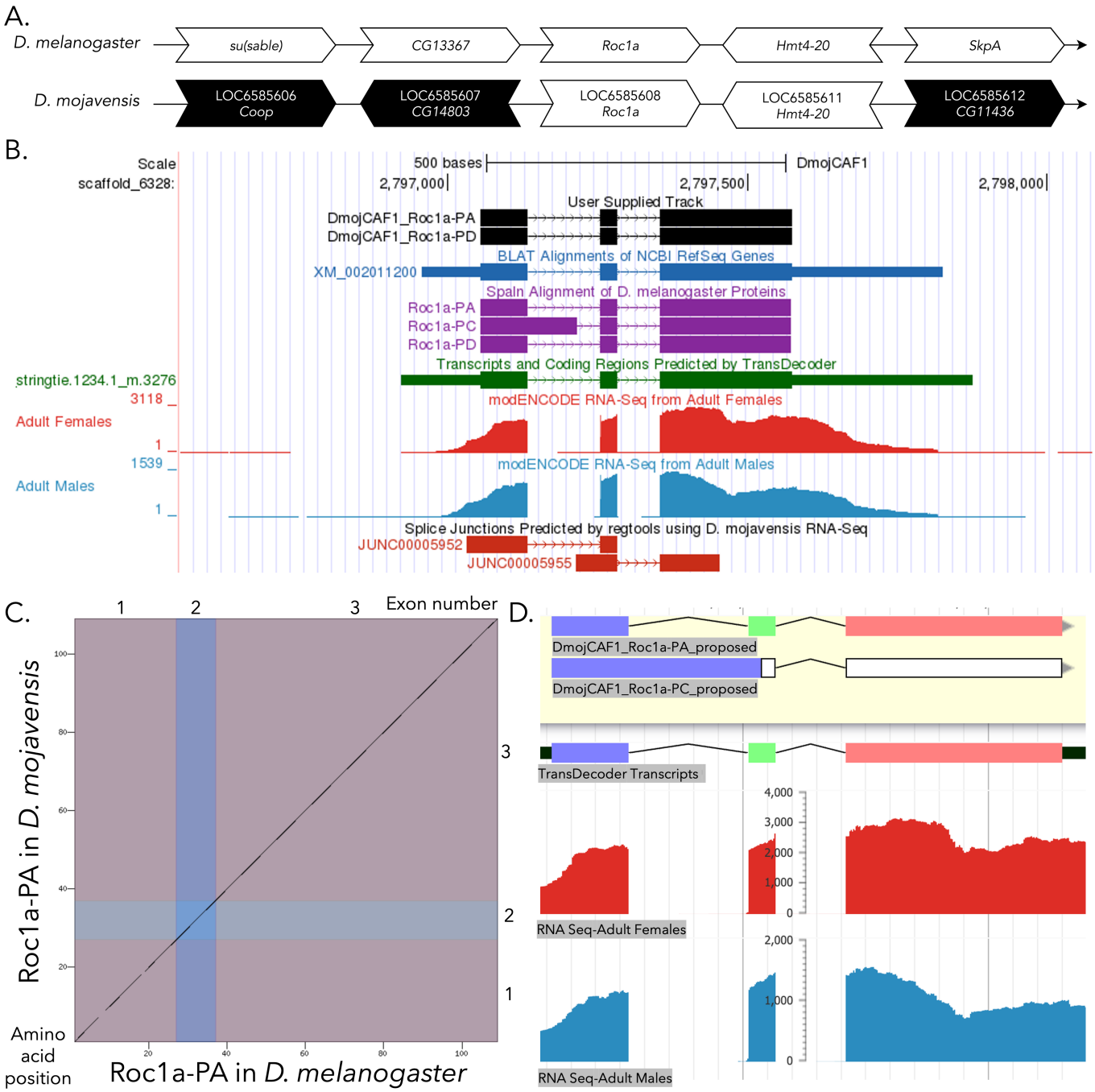

Figure 1. *Roc1a* gene model comparison between *Drosophila mojavensis* and *Drosophila melanogaster* orthologs:

(A) Synteny comparison of the genomic neighborhoods for *Roc1a* in *Drosophila melanogaster* and *D. mojavensis*.Thin underlying arrows indicate the DNA strand within which the reference gene–*Roc1a*–is located in *D. melanogaster* (top) and *D. mojavensis* (bottom) genomes. The thin arrows pointing to the right indicate that *Roc1a* is on the positive (+) strand in *D. melanogaster* and *D. mojavensis*. 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. mojavensis* indicate orthology to the corresponding gene in *D. melanogaster*, while black gene arrows indicate non-orthology. Gene symbols given in the *D. mojavensis* gene arrows indicate the orthologous gene in *D. melanogaster*, while the locus identifiers are specific to *D. mojavensis*. **(B) Gene Model in GEP UCSC Track Data Hub** (Raney et al., 2014). The coding-regions of *Roc1a* in *D. mojavensis* 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. mojavensis*), 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 and Adult Males (red and light blue, respectively; alignment of Illumina RNA-Seq reads from *D. mojavensis*), and Splice Junctions Predicted by regtools using *D. mojavensis* RNA-Seq (SRP006203). Splice junctions shown have a minimum read-depth of 100 with >1000 supporting reads shown in red. **(C) Dot Plot of Roc1a-PA in *D. melanogaster* (*x*-axis) vs. the orthologous peptide in *D. mojavensis* (*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. **(D) Model of Roc1a-PA compared to Roc1a-PC in Apollo.** A screenshot of the Apollo instance housing the proposed gene model of Roc1a-PA (identical to that of Roc1a-PD) and the likely absent Roc1a-PC model, containing in frame stop codons. The proposed models are shown at the top in the yellow region, while evidence tracks are found below in the white region. CDS reading frames are indicated in blue, green, and red, and implausible CDS sequence warnings (in this case due to the in-frame stop codons) are indicated in white. Evidence tracks from top to bottom include Transcripts and Coding Regions Predicted by TransDecoder and RNA-Seq from Adult Females and Adult Males (red and light blue, respectively; alignment of Illumina RNA-Seq reads from *D. mojavensis*; SRP006203).

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 (Noureddine 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. mojavensis (*NCBI:txid7230) is part of the *mulleri complex* in the *repleta* species group within the subgenus *Drosophila* of the genus *Drosophila* (Wasserman 1992; Durando et al., 2000). It was first described by Patterson (Patterson and Crow 1940). *D. mojavensis* specializes on rotting cactus as its host and is found in the Mojave and Sonoran Deserts of the southwestern United States and northwestern Mexico including the Baja Peninsula, as well as on the channel-islands off the coast of California (https://www.taxodros.uzh.ch, accessed 1 Feb 2023)." (Congleton et al., 2022).

We propose a gene model for the *D. mojavensis* ortholog of the *D. melanogaster Regulator of cullins 1a* (*Roc1a*) gene. The genomic region of the ortholog corresponds to the uncharacterized protein LOC6585608 (RefSeq accession XP_002011236.1) in the May 2011 (Agencourt dmoj_caf1/DmojCAF1) Genome Assembly of *D. mojavensis* (GenBank Accession: GCA_000005175.1). This model is based on RNA-Seq data from *D. mojavensis* (SRP006203 - Chen et al., 2014) 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-PA (query) against the *D. mojavensis* (GenBank Accession: GCA_000005175.1) Genome Assembly (database) placed the putative ortholog of *Roc1a* within scaffold_6328 (CH933812.1) at locus LOC6585608 (XP_002011236.1)— with an E-value of 1e-45 and a percent identity of 90.59%. Furthermore, the putative ortholog is flanked upstream by LOC6585607 (XP_002011235.1) and LOC6585606 (XP_002011234.1), which correspond to *CG14803* and *Coop* in *D. melanogaster* (E-value: 0.0 and .001; identity: 63.14% and 32.61%, respectively, as determined by *blastp*; Figure 1A, Altschul et al., 1990). The putative ortholog of *Roc1a* is flanked downstream by LOC6585611 (XP_043867559.1) and LOC6585612 (XP_032588764.1), which correspond to *Hmt4-20* and *CG11436* in *D. melanogaster* (E-value: 0.0 and 0.0; identity: 54.96% and 81.10%, respectively, as determined by *blastp*). The putative ortholog assignment for *Roc1a* in *D. mojavensis* is supported by the following evidence: the synteny of this genomic neighborhood is partially conserved, and the *BLAST* results for the target gene are of very good quality indicating a good match between the query sequence and the sequence at this location in *D. mojavensis.*

Protein Model

Roc1a in *D. mojavensis* has one unique protein-coding isoform encoded by mRNA isoforms *Roc1a-RA* and *Roc1a-RD* that differ in their UTRs, and contain three CDSs (Figure 1B). Relative to the ortholog in *D. melanogaster*, the RNA CDS number is conserved for *Roc1a-RA* and *Roc1a-RD*. However, *D. melanogaster* also has a third isoform with two CDSs, *Roc1a-RC*, that is not present in *D. mojavensis* (see: "special characteristics"). The sequence of Roc1a-PA in *D. mojavensis* has 97.22% identity (E-value: 2e-76) with the protein-coding isoform Roc1a-PA in *D. melanogaster*, as determined by *blastp* (Figure 1C). Coordinates of this curated gene model for Roc1a-PD and Roc1a-PA are stored by NCBI at GenBank/BankIt (accession **BK064491 and BK064492)**. These data are also archived in the CaltechDATA repository (see "Extended Data" section below).

Special characteristics of the protein model

Absence of Roc1a-PC in *D. mojavensis*

Isoform Roc1a-PC in *D. melanogaster* is similar to the other two isoforms, Roc1a-PA and Roc1a-PD. The difference between these mRNA isoforms is that *Roc1a-RA* and *Roc1a-RD* have three CDSs whereas *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; Larkin et al., 2021). In *D. melanogaster*, there are no in-frame stop codons for any of the isoforms, including in the longer first CDS of *Roc1a-RC*. However, in *D. mojavensis*, there is an in-frame stop codon present in the first CDS of Roc1a-RC that prematurely stops translation of this isoform. This has been highlighted in Figure 1D by the white portion of the CDSs of *Roc1a-RC* indicating the presence of in-frame stop codons. This, in combination with the lack of RNA-seq data and TransDecoder Transcript predictions supporting the existence of Roc1a-PC, suggests that Roc1a-PC is likely absent from *D. mojavensis* (Figure 1D). This special characteristic of Roc1a-PC is similar to the observations described in Lawson et al. (2025).

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. melanogaste*r 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: DmojCAF1_Roc1a.zip. DOI: 10.22002/t53jr-79t54

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

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

Congleton H, Kiser CA, Colom Diaz PA, Schlichting E, Walton DA, Long LJ, et al., Rele CP. 2022. Drosophila mojavensis - chico. MicroPubl Biol 2022. PubMed ID: 36468157

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

Durando CM, Baker RH, Etges WJ, Heed WB, Wasserman M, DeSalle R. 2000. Phylogenetic analysis of the repleta species group of the genus Drosophila using multiple sources of characters. Mol Phylogenet Evol 16(2): 296-307. PubMed ID: 10942616

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: 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

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

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 I, Alvarado B, German T, Thompson JS, Long LJ, DiAngelo JR, Yang M, Rele CP. 2025. Gene model for the ortholog of Roc1a in Drosophila eugracilis, microPublication Biology, submitted 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

Patterson JT and JF Crow, 1940. Hybridization in the mulleri group of Drosophila. Univ. Texas Publs, 4032, 167-189 PubMed ID: null

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

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

Wasserman, M. (1992). Cytological evolution of the Drosophila repleta species group. Krimbas, Powell, 1992 : 455-552. FBrf0063954 PubMed ID: null

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; https://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. Ryan J. Dufur: formal analysis, writing - review editing. Braden Wright: formal analysis, writing - review editing. Isabel G. Wellik: formal analysis, writing - review editing. Lindsey J. Long: supervision, writing - review editing. Jeffrey S. Thompson: 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, funding acquisition, conceptualization, project administration, writing - review editing.

Reviewed By: Anonymous

Nomenclature Validated By: Anonymous

History: Received October 21, 2023 **Revision Received** February 12, 2025 **Accepted** February 4, 2025 **Published Online** February 13, 2025 **Indexed** February 27, 2025

Citation: Lawson, ME; Dufur, RJ; Wright, B; Wellik, IG; Long, LJ; Thompson, JS; Rele, CP; Reed, LK (2025). Gene model for the ortholog of *Roc1a* in *Drosophila mojavensis*. microPublication Biology. 10.17912/micropub.biology.001038