microPublication BIOLOGY

11/12/2024 - *Open Access*

Gene model for the ortholog of *DENR* in *Drosophila yakuba*

Leon F. Laskowski[1], Inayah Burton[2], Timothy J. Stanek[3], Geoffrey D. Findlay[4], Scott Tanner[5], Jack A. Vincent[6], Solomon Tin Chi Chak[2], Christopher E. Ellison[3], Chinmay P. Rele[1§]

[1]University of Alabama, Tuscaloosa, Alabama, USA

[2]SUNY Old Westbury, Old Westbury, NY USA

[3]Rutgers University, New Brunswick, NJ USA

[4]College of the Holy Cross, Worcester, MA USA

[5]University of South Carolina Upstate, Spartanburg, SC USA

[6]University of Washington - Tacoma, Tacoma, WA USA

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

Abstract

Gene model for the ortholog of *Density regulated protein* (*DENR*) in the May 2011 (WUGSC dyak_caf1/DyakCAF1) Genome Assembly (GenBank Accession: GCA_000005975.1) of *Drosophila yakuba*. 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.

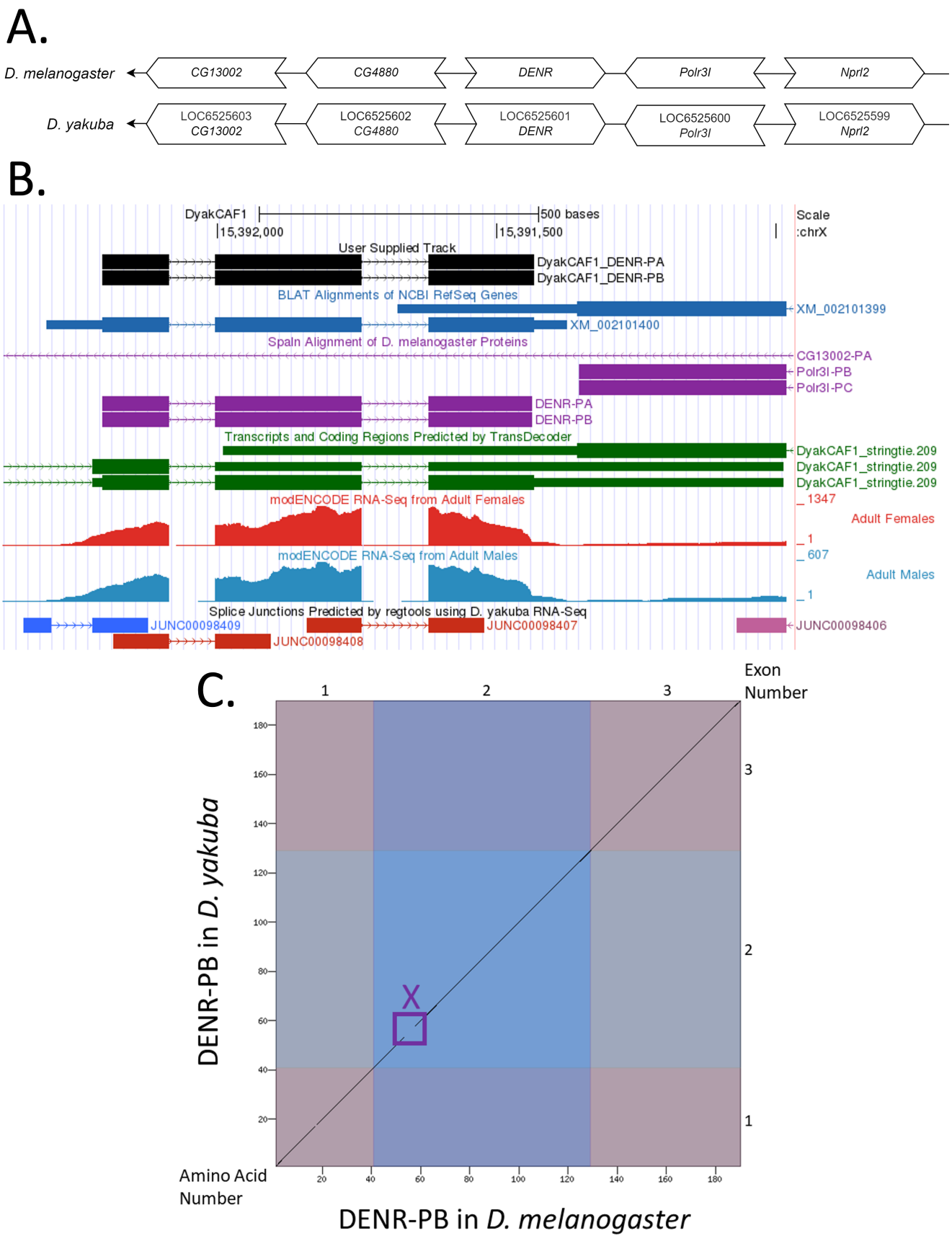
A.
D. melanogaster
CG13002
CG4880
DENR
Polr3I
Nprl2
D. yakuba
LOC6525603 CG13002
LOC6525602 CG4880
LOC6525601 DENR
LOC6525600 Polr3I
LOC6525599 Nprl2
B.
DyakCAF1
500 bases
15,392,000
15,391,500
Scale
chrX
User Supplied Track
DyakCAF1_DENR-PA
DyakCAF1_DENR-PB
BLAT Alignments of NCBI RefSeq Genes
XM_002101399
XM_002101400
Spaln Alignment of D. melanogaster Proteins
CG13002-PA
Polr3I-PB
Polr3I-PC
DENR-PA
DENR-PB
Transcripts and Coding Regions Predicted by TransDecoder
DyakCAF1_stringtie.209
DyakCAF1_stringtie.209
DyakCAF1_stringtie.209
modENCODE RNA-Seq from Adult Females
_1347
Adult Females
_1
modENCODE RNA-Seq from Adult Males
_607
Adult Males
_1
Splice Junctions Predicted by regtools using D. yakuba RNA-Seq
JUNC00098409
JUNC00098407
JUNC00098406
JUNC00098408
C.
Exon Number
1
2
3
DENR-PB in D. yakuba
DENR-PB in D. melanogaster
Amino Acid Number
20
40
60
80
100
120
140
160
180

Figure 1.

(A) Synteny comparison of the genomic neighborhoods for *DENR* in *Drosophila melanogaster* and *D. yakuba*. Thin underlying arrows indicate the DNA strand within which the target gene–*DENR*–is located in *D. melanogaster* (top) and *D. yakuba* (bottom). The thin arrows pointing to the left indicate that *DENR* is on the negative (-) strand in *D. yakuba* and *D. melanogaster*. The wide gene arrows pointing in the same direction as *DENR* are on the same strand relative to the thin underlying arrows, while wide gene arrows pointing in the opposite direction of *DENR* are on the opposite strand relative to the thin underlying arrows. White gene arrows in *D. yakuba* indicate orthology to the corresponding gene in *D. melanogaster*. Gene symbols given in the *D. yakuba* gene arrows indicate the orthologous gene in *D. melanogaster*, while the locus identifiers are specific to *D. yakuba*. **(B) Gene Model in GEP UCSC Track Data Hub (Raney et al., 2014).** The coding-regions of *DENR* in *D. yakuba* are displayed in the User Supplied Track (black); 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. yakuba*), 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. yakuba*), and Splice Junctions Predicted by regtools using *D. yakuba* RNA-Seq (SRP006203). Splice junctions shown in blue (JUNC00098409) or red (JUNC00098408; JUNC00098407) have read-depths of 21, 1039 and 1780, respectively. **(C) Dot Plot of DENR-PB in *D. melanogaster* (*x*-axis) vs. the orthologous peptide in *D. yakuba* (*y*-axis).** Amino acid number is indicated along the left and bottom; CDS number is indicated along the top and right, CDSs are also highlighted with alternating colors. The purple box, X, indicates a lack of sequence similarity between amino acids.

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 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. yakuba* (NCBI:txid 7245) 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 Burla (1954). *D. yakuba* is wide-spread in sub-Saharan Africa and Madagascar (Lemeunier et al., 1986; https://www.taxodros.uzh.ch, accessed 1 Feb 2023; Markow and O'Grady 2006) where figs served as a primary host along with other rotting fruits (Lachaise and Tsacas 1983)." (Koehler et al., 2024).

We propose a gene model for the *D. yakuba* ortholog of the *D. melanogaster Density regulated protein* (*DENR*) gene. The genomic region of the ortholog corresponds to the uncharacterized protein LOC6525601 (RefSeq accession XP_002101436.1) in the Dyak_CAF1 Genome Assembly of *D. yakuba* (GenBank Accession: GCA_000005975.1 - Graveley et al., 2011). This model is based on RNA-Seq data from *D. yakuba* (SRP006203) and *DENR* in *D. melanogaster* using FlyBase release FB2022_04 (GCA_000001215.4; Larkin et al., 2021).

Density regulated protein was first discovered in a human teratocarcinoma cell line because its concentration in cells increased with cell density (Deyo et al., 1998). Subsequent bioinformatic and biochemical analyses showed that the protein is conserved across eukaryotes and functions in non-canonical translation initiation (Fleischer et al., 2006; Skabkin et al., 2010). *D. melanogaster* flies homozygous for a null, knockout allele of the gene encoding *Density regulated protein*, *DENR*

(FBgn0030802), die as pharate adults, showing a larval-like epidermis and reduced proliferation of histoblast cells (Schleich et al., 2014). Subsequent experiments using both RNAi in S2 cells and the knockout allele in larvae showed that DENR is required, along with its interacting partner MCT-1, for the proper expression regulation of a subset of transcripts required for cell cycle progression and growth. In particular, the loss of *DENR* reduces expression of the insulin receptor and makes larvae less sensitive to insulin signaling (Schleich et al., 2014), thus implicating DENR in the regulation of the insulin signaling pathway.

Synteny

The target gene, *DENR*, occurs on chromosome X in *D. melanogaster* and is flanked upstream by *CG13002* and *CG4880* and downstream by *RNA polymerase III subunit I* (*Polr3I*) and *Nitrogen permease regulator-like 2* (*Nprl2*). The *tblastn* search of *D. melanogaster* DENR-PB (query) against the *D. yakuba* (GenBank Accession: GCA_000005975.1) Genome Assembly (database) placed the putative ortholog of *DENR* within scaffold chromosome X (CM000162.2) at locus LOC6525601 (XP_002101436.1)— with an E-value of 3e-51 and 59.66% identity. Furthermore, the putative ortholog is flanked upstream by LOC6525603 (XP_002101438.1) and LOC6525602 (XP_043062942.1), which correspond to *CG13002* and *CG4880* in *D. melanogaster* (E-value: 2e-95 and 1e-150; identity: 67.53% and 71.18%, respectively, as determined by *blastp*; Figure 1A, Altschul et al., 1990). The putative ortholog of *DENR* is flanked downstream by LOC6525600 (XP_002101435.1) and LOC6525599 (XP_002101434.1), which correspond to *Polr3I* and *Nprl2* in *D. melanogaster* (E-value: 7e-94 and 0.0; identity: 85.54% and 99.73%, respectively, as determined by *blastp*). The putative ortholog assignment for *DENR* in *D. yakuba* is supported by the following evidence: The genes surrounding the *DENR* ortholog are orthologous to the genes at the same locus in *D. melanogaster* and local synteny is completely conserved, supported by results generated from *blastp*, so we conclude that LOC6525601 is the correct ortholog of *DENR* in *D. yakuba* (Figure 1A).

Protein Model

Based on the annotation of two mRNA isoforms with identical protein-coding sequences in *D. melanogaster*, *DENR* in *D. yakuba* is predicted to have one unique protein-coding isoform (*DENR-PB* and *DENR-PA*; Figure 1B). mRNA isoforms (*DENR-RB*; *DENR-RA*) contain three CDS each. Relative to the ortholog in *D. melanogaster*, the RNA CDS number and protein isoform count are conserved. The sequence of DENR-PB in *D. yakuba* has 98.41% identity (E-value: 7e-97) with the protein-coding isoform DENR-PB in *D. melanogaster*, as determined by *blastp* (Figure 1C). Box X in purple highlights a gap in the dot plot, indicating a lack of sequence similarity in that region (Figure 1C). Coordinates of this curated gene model of DENR-PA and DENR-PB are stored by NCBI at GenBank/BankIt (accession BK064477 and BK064478, respectively**)**. These data are also archived in the CaltechDATA repository (see “Extended Data” section below).

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* target gene for a given isoform and run it using *tblastn* against their target *Drosophila* species genome assembly (*Drosophila yakuba* (GCA_000005975.1)- Graveley et al., 2010)) 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. Genomic structure information (e.g., CDSs, CDS 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 this gene were generated by students under mentorship of their faculty course instructors. These models were then reconciled by a third independent researcher mentored by the project leaders to produce

the final model presented here. 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 and Laura K. Reed for overseeing the project. Thank you to FlyBase for providing the definitive database for *Drosophila melanogaster* gene models. FlyBase is supported by grants: NHGRI U41HG000739 and U24HG010859, UK Medical Research Council MR/W024233/1, NSF 2035515 and 2039324, BBSRC BB/T014008/1, and Wellcome Trust PLM13398.

Extended Data

Description: A GFF, FASTA, and PEP of the model. Resource Type: Model. File: DyakCAF1_DENR.zip. DOI: 10.22002/bg0b6-1z133

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.

Burla H. 1954. Zur Kenntnis der Drosophiliden der Elfenbeinkuste (Franzosisch West-Afrika). Revue suisse Zool. 61(Suppl.): 1--218. FBrf0009861.

Deyo JE, Chiao PJ, Tainsky MA. 1998. drp, a Novel Protein Expressed at High Cell Density but Not During Growth Arrest. DNA and Cell Biology 17: 437-447. DOI: 10.1089/dna.1998.17.437

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., the FlyBase Consortium. 2022. FlyBase: a guided tour of highlighted features. Genetics 220(4). 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

Fleischer TC, Weaver CM, McAfee KJ, Jennings JL, Link AJ. 2006. Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. Genes Dev 20(10): 1294-307. PubMed ID: 16702403

Hietakangas V, Cohen SM. 2009. Regulation of tissue growth through nutrient sensing. Annu Rev Genet 43: 389-410. PubMed ID: 19694515

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

Koehler AC, Romo I, Le V, Romo I, Youngblom JJ, Hark AT, Rele CP. 2024. Gene model for the ortholog of *Glys* in *Drosophila yakuba*. microPublication Biology. (submitted)

Lachaise D, Tsacas L. 1983. Breeding-sites of tropical African Drosophilids. Ashburner, Carson, Thompson, 1981-1986. 3d: 21--332. FBrf0038884.

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

Lemeunier F, David J, Tsacas L, Ashburner M. 1986. The melanogaster species group. Ashburner, Carson, Thompson, 1981-1986. e: 147--256. FBrf0043749.

Markow TA, O'Grady P. 2005. Drosophila: A guide to species identification and use. Academic Press 978-0-12-473052-6 PubMed ID: 5073854

Mudge JM, Harrow J. 2016. The state of play in higher eukaryote gene annotation. Nat Rev Genet 17(12): 758-772. PubMed ID: 27773922

11/12/2024 - *Open Access*

Myers A, Hoffmann A, Natysin M, Arsham AM, Stamm J, Thompson JS, Rele CP 2024. Gene model for the ortholog *Myc* in *Drosophila ananassae*, microPublication Biology. (submitted)

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. 2022. Manual annotation of Drosophila genes: a Genomics Education Partnership protocol. F1000Res 11: 1579. PubMed ID: 37854289

Schleich S, Strassburger K, Janiesch PC, Koledachkina T, Miller KK, Haneke K, et al., Teleman AA. 2014. DENR-MCT-1 promotes translation re-initiation downstream of uORFs to control tissue growth. Nature 512(7513): 208-212. PubMed ID: 25043021

Skabkin MA, Skabkina OV, Dhote V, Komar AA, Hellen CU, Pestova TV. 2010. Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. Genes Dev 24(16): 1787-801. PubMed ID: 20713520

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 (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.

Author Contributions: Leon F. Laskowski: formal analysis, validation, writing - original draft, writing - review editing. Inayah Burton: formal analysis, writing - review editing. Timothy J. Stanek: formal analysis, writing - review editing. Geoffrey D. Findlay: writing - original draft, writing - review editing. Scott Tanner: writing - original draft, writing - review editing. Jack A. Vincent: writing - original draft, writing - review editing. Solomon Tin Chi Chak: supervision, writing - review editing. Christopher E. Ellison: supervision, writing - review editing. Chinmay P. Rele: data curation, formal analysis, methodology, project administration, software, supervision, validation, visualization, writing - review editing.

Reviewed By: David Molik

History: Received October 4, 2023 **Revision Received** June 30, 2024 **Accepted** October 21, 2024 **Published Online** November 12, 2024 **Indexed** November 26, 2024

Citation: Laskowski, LF; Burton, I; Stanek, TJ; Findlay, GD; Tanner, S; Vincent, JA; et al.; Rele, CP (2024). Gene model for the ortholog of *DENR* in *Drosophila yakuba*. microPublication Biology. 10.17912/micropub.biology.001017