

Gene model for the ortholog of *S6k* in *Drosophila ananassae*

Anne E. Backlund¹, Joyce Lai², Jordan Hensley³, Anya Goodman², James J. Youngblom³, Chinmay P. Rele¹, Laura K Reed^{1§}

¹The University of Alabama, Tuscaloosa, AL USA
²California Polytechnic State University, San Luis Obispo, CA USA
³California State University Stanislaus, Turlock, CA USA
[§]To whom correspondence should be addressed: lreed1@ua.edu

Abstract

Gene model for the ortholog of *Ribosomal protein S6 kinase (S6k)* in the May 2011 (Agencourt dana_caf1/DanaCAF1) Genome Assembly (GenBank Accession: 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.

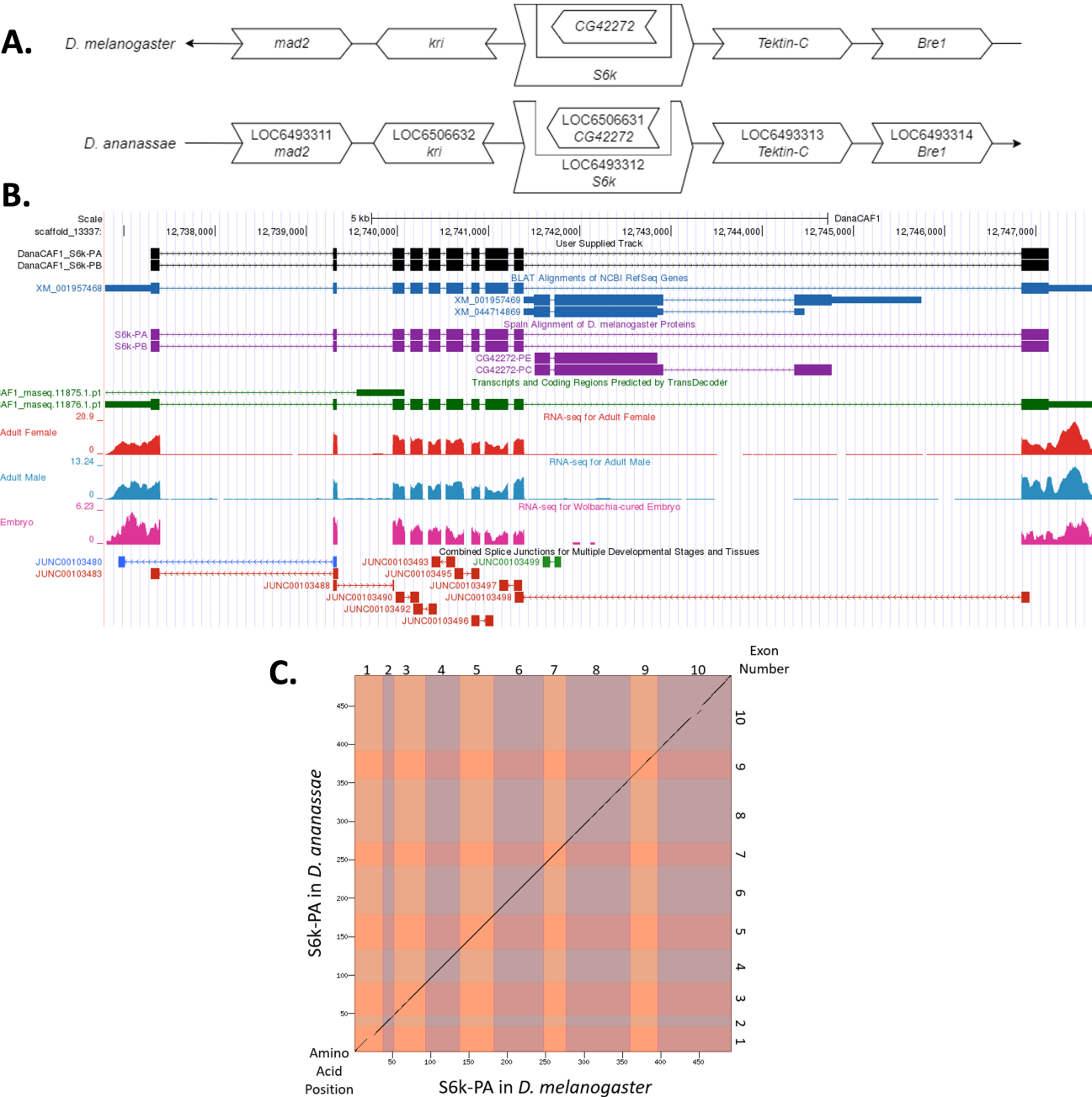


Figure 1. *S6k* gene model comparisons between *Drosophila ananassae* and *Drosophila melanogaster* orthologs:

(A) Synteny comparison of the genomic neighborhoods for *S6k* in *Drosophila melanogaster* and *D. ananassae*. Thin underlying arrows indicate the DNA strand within which the target gene—*S6k*—is located in *D. melanogaster* (top) and *D. ananassae* (bottom). The thin arrow pointing to the right indicates that *S6k* is on the positive (+) strand in *D. ananassae*, and the thin arrow pointing to the left indicates that *S6k* is on the negative (-) strand in *D. melanogaster*. The wide gene arrows pointing in the same direction as *S6k* are on the same strand relative to the thin underlying arrows, while wide gene arrows pointing in the opposite direction of *S6k* 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 *S6k* 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*), 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. ananassae*), and Splice Junctions Predicted by regtools using *D. ananassae* RNA-Seq (SRP006203, SRP007906, PRJNA257286, PRJNA388952). Splice junctions shown have a minimum read-depth of 10 with 10-49, 50-99, >1000 supporting reads in blue, green, and red, respectively **(C) Dot Plot of *S6k*-PA 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.

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).

"Ribosomal protein S6 kinase ([S6k](#) aka *p70S6K*, FBgn0283472) is part of the insulin signaling pathway downstream of the target of rapamycin (*dTOR*) (Toker 2000), homologous to mammalian *p70S6k* (Watson et al., 1996). *S6k* is a serine/threonine kinase in *Drosophila melanogaster* and acts as a regulator of cell size (Montagne et al., 1999), as well as innate immunity and senescence (Fabian et al., 2021)." (Keirn et al., 2024).

"*D. ananassae* (NCBI:txid 7217) 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* Ribosomal protein S6 kinase (*S6k*) gene. The genomic region of the ortholog corresponds to the uncharacterized protein [LOC6493312](#) (RefSeq accession [XP_001957504.1](#)) in the May 2011 (Agencourt dana_caf1/DanaCAF1) Genome Assembly of *D. ananassae* ([GCA_000005115.1](#); Drosophila 12 Genomes Consortium et al., 2007). This model is based on RNA-Seq data from *D. ananassae* ([SRP006203](#), [SRP007906](#); [PRJNA257286](#), [PRJNA388952](#); Graveley et al., 2011) and *S6k* in *D. melanogaster* using FlyBase release FB2023_02 ([GCA_000001215.4](#); Larkin et al., 2021; Gramates et al., 2022; Jenkins et al., 2022).

Syteny

The reference gene, *S6k*, occurs on chromosome 3L in *D. melanogaster* and is flanked upstream by *mad2* and *krishah* (*kri*) and downstream by *Tektin-C* and *Bre1*. The *tblastn* search of *D. melanogaster* *S6k*-PA (query) against the *D. ananassae* (GCA_000005115.1) Genome Assembly (database) placed the putative ortholog of *S6k* within scaffold_13337 at locus [LOC6493312 \(XP_001957504.1\)](#)— with an E-value of 8e-138 and a percent identity of 56.32%. Furthermore, the putative ortholog is flanked upstream by [LOC6493311 \(XP_001957502.1\)](#) and [LOC6506632 \(XP_032310700.1\)](#), which correspond to *mad2* and *kri* in *D. melanogaster* (E-value: 1e-149 and 1e-158; percent identity: 95.17% and 94.25%, respectively, as determined by *blastp*; Figure 1A, Altschul et al., 1990). The putative ortholog nests another gene, [LOC6506631 \(XP_001957505.1\)](#), which corresponds to [CG42272](#) in *D. melanogaster* (E-value: 0.0, percent identity: 76.51%). The putative ortholog of *S6k* is flanked downstream by [LOC6493313 \(XP_001957506.1\)](#) and [LOC6493314 \(XP_014764325.1\)](#), which correspond to *Tektin-C* and *Bre1* in *D. melanogaster* (E-value: 0.0 and 0.0; percent identity: 98.81% and 92.62%, respectively, as determined by *blastp*). The putative ortholog assignment for *S6k* in *D. ananassae* is supported by the following evidence: The genomic neighborhood is conserved in *D. ananassae* and the *blastp* results for the putative ortholog of *S6k* had a low E-value and a high percent identity. We conclude that this is the location of the ortholog of *S6k* in *D. ananassae*.

Protein Model

S6k in *D. ananassae* has one unique protein-coding isoform (*S6k*-PA and *S6k*-PB; Figure 1B), encoded by mRNA isoforms (*S6k*-RA and *S6k*-RB that differ in their UTRs) that contain ten CDSs. Relative to the ortholog in *D. melanogaster*, the CDS number is conserved. The sequence of *S6k*-PA in *D. ananassae* has 95.36% identity (E-value: 0.0) with the protein-coding isoform *S6k*-PA in *D. melanogaster*, as determined by *blastp* (Figure 1C). Coordinates of this curated gene model (*S6k*-PB, *S6k*-PA) are stored by NCBI at GenBank/BankIt (accession **BK064576**, **BK064577**, 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* 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 syteny 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: This publication is dedicated to the memory of Dr. James J. Youngblom. 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: [DanaCAF1_S6k.zip](#). DOI: [10.22002/e86gh-a7p16](https://doi.org/10.22002/e86gh-a7p16)

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215: 403-410. PubMed ID: [2231712](#)
- Bock IR, Wheeler MR. 1972. The *Drosophila melanogaster* species-group. *Univ. Texas Publs Stud. Genet.* 7213: 1-102. FBrf0024428.
- Doleschall CL. 1858. Derde bijdrage tot de kennis der dipterologische fauna van Nederlandsch Indië. *Natuurkundig tijdschrift voor Nederlandsch Indië.* 17: 73. FBrf0000091
- 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](#)
- Fabian DK, Fuentealba M, Dönertaş HM, Partridge L, Thornton JM. 2021. Functional conservation in genes and pathways linking ageing and immunity. *Immun Ageing* 18(1): 23. PubMed ID: [33990202](#)
- 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](#)
- 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: 1006-1010. 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](#)
- 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](#)
- Keirn G, Kiser CA, Laskowski LF, Hensley J, Mortan R, Nguyen T, et al., Reed LK. 2024. Gene model for the ortholog of S6k in *Drosophila yakuba*. *MicroPubl Biol* 2024: 10.17912/micropub.biology.001018. PubMed ID: [39758581](#)
- 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. FBrf0004281 DOI: [10.1007/bf01531779](#)
- 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, 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: 10.17912/micropub.biology.000782. PubMed ID: [39717145](#)
- Markow TA, O'Grady P. 2005. *Drosophila: A guide to species identification and use.* Academic Press 978-0-12-473052-6
- Montagne J, Stewart MJ, Stocker H, Hafen E, Kozma SC, Thomas G. 1999. *Drosophila* S6 kinase: a regulator of cell size. *Science.* 285: 2126-2129. PubMed ID: [10497130](#)
- Mudge JM, Harrow J. 2016. The state of play in higher eukaryote gene annotation. *Nat. Rev. Genet.* 17: 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. 2022. Manual annotation of *Drosophila* genes: a Genomics Education Partnership protocol. *F1000Res* 11: 1579. PubMed ID: [37854289](#)
- Singh BN. 2010. *Drosophila ananassae*: a good model species for genetical, behavioural and. *Indian J. Exp. Biol.* 48: 333-345. PubMed ID: [20726331](#)

9/11/2025 - Open Access

Singh BN, Yadav JP. 2015. Status of research on *Drosophila ananassae* at global level. J. Genet.. 94: 785-792. PubMed ID: [26690536](#)

Sturtevant AH. 1939. On the subdivision of the genus *Drosophila*. Proc. Natl. Acad. Sci. U. S. A.. 25: 137-141. 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](#)

Toker A. 2000. Protein kinases as mediators of phosphoinositide 3-kinase signaling. Mol Pharmacol. 57: 652-658. PubMed ID: [10727509](#)

Watson KL, Chou MM, Blenis J, Gelbart WM, Erikson RL. 1996. A *Drosophila* gene structurally and functionally homologous to the mammalian 70-kDa s6 kinase gene. Proc Natl Acad Sci U S A 93(24): 13694-8. PubMed ID: [8942996](#)

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

Supported by National Science Foundation (United States) 1915544 to LK Reed.

Author Contributions: Anne E. Backlund: formal analysis, validation, writing - original draft, writing - review editing. Joyce Lai: formal analysis, writing - review editing. Jordan Hensley: formal analysis, writing - review editing. Anya Goodman: supervision, writing - review editing. James J. Youngblom: 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, methodology, writing - original draft.

Reviewed By: Jacob Kagey

Nomenclature Validated By: Anonymous

History: Received October 11, 2023 **Revision Received** September 2, 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: Backlund AE, Lai J, Hensley J, Goodman A, Youngblom JJ, Rele CP, Reed LK. 2025. Gene model for the ortholog of *S6k* in *Drosophila ananassae*. microPublication Biology. [10.17912/micropub.biology.001030](#)