

Drosophila mojavensis – chico


Hannah Congleton¹, Cole A. Kiser², Patricia A. Colom Diaz³, Elizabeth Schlichting², Dorothy A. Walton¹, Lindsey J. Long¹, Laura K. Reed², Juan Carlos Martinez-Cruzado³, Chinmay P. Rele^{2§}

¹Oklahoma Christian University, Edmond, OK USA

²The University of Alabama, Tuscaloosa, AL USA

³University of Puerto Rico at Mayagüez, Mayagüez, PR Puerto Rico

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

 **Figure 1.**

(A) Synteny comparison of the genomic neighborhoods for *chico* in *Drosophila melanogaster* and *D. mojavensis*. Thin underlying arrows indicate the DNA strand within which the target gene—*chico*—is located on the negative strand in *D. melanogaster* (top) and *D. mojavensis* (bottom). The wide gene arrows pointing in the same direction as *chico* are on the same strand relative to the thin underlying arrows, while wide gene arrows pointing in the opposite direction of *chico* 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 *chico* in *D. mojavensis* are displayed in the User Supplied Track (black); coding exons 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 (Chen et al., 2014; SRP006203). Splice junctions shown have a read-depth of 10-49 and 100-499 supporting reads in blue and pink. **(C) Dot Plot of *chico*-PA in *D. melanogaster* (x-axis) vs. the new PAA isoform in *D. mojavensis* (y-axis).** Amino acid number is indicated along the left and bottom; coding-exon number is indicated along the top and right, and exons are also highlighted with alternating colors. The red circle labeled 1, Regions which indicate a lack of similarity between the two sequences are highlighted in red (Region 1; Region 2). **(D) Gene model in GEP UCSC displaying RNA-Seq and Splice Junctions data of exons six and seven of the ortholog in *D. mojavensis*.** Boxes A and B indicate RNA-Seq drop-offs, boxes C and D indicate their corresponding splice acceptors and the region in between the boxes represent introns. The increase of RNA-Seq data in exons 6 and 7 and the existence of splice junction JUNC00059058 support that an additional isoform is present in *D. mojavensis* (*chico*-PAA).

Description

IntroductionThe insulin signaling pathway is a highly conserved pathway in animals and is central to nutrient uptake (Hietakangas and Cohen 2009; Grewal 2009). A key component of the pathway, encoding a substrate of the Insulin Receptor (InR) protein, *chico* binds to the InR and thus is also referred to as the Insulin Receptor Substrate (IRS). The IRS *chico* is essential to the control of cell and organ size (Goderdhan et al. 1999; Bohni et al. 1999), overall growth (Bakopoulos et al. 2020), including the sex-biased control of cell size and growth (Millington et al. 2021; Kim and O'Connor 2021), and is involved in lifespan (Clancy et al. 2001). We propose a gene model for the *D. mojavensis* ortholog of the *D. melanogaster* *chico* (*chico*) gene. The genomic region of the ortholog corresponds to the uncharacterized protein LOC6576893 (RefSeq accession XP_032585220.1) in the dmoj_caf1 Genome Assembly of *D. mojavensis* (GenBank Accession: GCA_000005175.1 - Chen et al., 2014; SRP006203). This model is based on RNA-Seq data from *D. mojavensis* (Chen et al., 2014; SRP006203) and *chico* in *D. melanogaster* using FlyBase release FB2022_04 (GCA_000001215.4; Larkin et al., 2021). *D. mojavensis* 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>). The Genomics Education Partnership maintains a mirror of the UCSC Genome Browser (Kent WJ et al., 2002; Gonzalez et al., 2021), which is available at <https://gander.wustl.edu>.

Synteny

The target gene, *chico*, occurs on chromosome 2L in *D. melanogaster* and is flanked upstream by *CG31717* and basket (*bsk*) and downstream by maternal expression at 31B (*me31B*) and methuselah-like 15 (*mthl15*). The *tblastn* search of *D. melanogaster* *chico*-PA (query) against the *D. mojavensis* (GenBank Accession: GCA_000005175.1) Genome Assembly (database) placed the putative ortholog of *chico* within scaffold scaffold_6500 (CH933807.1) at locus LOC6576893 (XP_032585220.1)— with an E-value of 0.0 and a percent identity of 44.97%. Furthermore, the putative ortholog is flanked upstream by LOC6576896 (XP_015020894.1) LOC6576895 (XP_002002875.4) and LOC6576894 (XP_015020893.1), which correspond to *CG31717*, Multi-substrate lipid kinase (*Mulk*) and trunk (*trk*) in *D. melanogaster* (E-value: 2e-77, 3e-155 and 3e-100; identity: 65.64%, 54.44% and 65.58%, respectively, as determined by *blastp*; Figure 1A, Altschul *et al.*, 1990). The putative ortholog of *chico* is flanked downstream by LOC6576891 (XP_002002871.1) and LOC6576890 (XP_002002870.1), which correspond to *CG33303* and *CG5731* in *D. melanogaster* (E-value: 0.0 and 0.0; identity: 65.78% and 84.56%, respectively, as determined by *blastp*). The putative ortholog assignment for *chico* in *D. mojavensis* is supported by the following evidence: Although the genes surrounding the *chico* ortholog are not orthologous to the genes at the same locus in *D. melanogaster*, we conclude that LOC6576893 is the correct ortholog of *chico* in *D. mojavensis* (Figure 1A), supported by strong e-values and percent identities and the lack of other matches while performing the *blastp* search.

Protein Model

Consistent with the *blastp* search result which shows 44.97% identity between *D. melanogaster* *chico*-PA and the *D. mojavensis* gene model as well as the low sensitivity parameters used to generate the dot plot (i.e., word size = 3; neighborhood threshold = 11), the dot plot of the two protein sequences contain multiple large gaps along the diagonal. *chico* in *D. mojavensis* has 3 identical protein-coding isoforms (*chico*-PA, *chico*-PB and *chico*-PC; Figure 1B). Isoform (*chico*-PA, *chico*-PB, *chico*-PC) contains 8 protein-coding exons. Relative to the ortholog in *D. melanogaster*, the coding-exon number is conserved. The sequence of *chico*-PA in *D. mojavensis* has 44.97% identity (E-value: 0.0) with the protein-coding isoform *chico*-PA in *D. melanogaster*, as determined by *blastp* (Figure 1C). While conducting our research, the presence of a new isoform of *chico* in *D. mojavensis* was discovered. The isoform identified as PA *D. mojavensis* is orthologous to the PA isoform in *D. melanogaster* and the PAA isoform is the novel isoform. Lower score Splice Junctions data due to lower levels of RNA-Seq data indicate the *chico*-PAA isoform is expressed at lower levels, or in fewer cells, than the *chico*-PA isoform (Figure 1D). Coordinates of this curated gene model are stored by NCBI at GenBank/BankIt (accession BK059533). These data are also archived in the CaltechDATA repository (see “Extended Data” section below).

Special characteristics of the protein model

Novel isoform: A new isoform of *chico* (*chico*-PAA) has been discovered based on the available RNA-Seq data and Splice Junctions data (Figure 1D). RNA-Seq data drops off in two regions, labeled A and B, and RNA-Seq data increase are displayed in two regions, labeled C and D. Blue and pink data tracks beneath the RNA-Seq data represent Splice Junctions, with pink splice junctions having a higher score than blue splice junctions (minimum read-depth scores found in above image caption of Figure 1A). The existence of splice junction JUNC00059058 as well as the decrease and sharp increase in RNA-Seq data in exons 6 and 7 which overlap the exons of isoform PA indicates an additional isoform (Figure 1D).

Acknowledgements: We would like to thank Wilson Leung for developing and maintaining the technological infrastructure that was used to create this gene model, and Madeline Gruys for retrofitting this model.

Extended Data

Description: FASTA. Resource Type: Model. File: [DmojCAF1_chico.fasta](#). DOI: [10.22002/n7ehe-xej65](#)

Description: GFF. Resource Type: Model. File: [DmojCAF1_chico.gff](#). DOI: [10.22002/26d2v-g8r63](#)

Description: Peptide Sequence. Resource Type: Model. File: [DmojCAF1_chico.pep](#). DOI: [10.22002/vbfq6-4z544](#)

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). "Basic local alignment search tool." *J. Mol. Biol.* 215:403-410. PMID: 2231712 PubMed ID: [2231712](#)
- Bakopoulos, D., Beadle, L.F., Esposito, K.M., Mirth, C.K., Warr, C.G., Johnson, T.K. 2020 Insulin-Like Signalling Influences the Coordination of Larval Hemocyte Number with Body Size in *Drosophila melanogaster*. FBrf0246140 PMID: 32341056 PubMed ID: [32341056](#)
- Bohni R, Riesgo-Escovar J, Oldham S, Brogiolo W, Stocker H, Andruss BF, Beckingham K, Hafen E. 1999. Autonomous Control of Cell and Organ Size by CHICO, a *Drosophila* Homolog of Vertebrate IRS1-4. *Cell.* 97(7): 865–875. PMID: 10399915 PubMed ID: [10399915](#)

Chen ZX, Sturgill D, Qu J, Jiang H, Park S, Boley N, Suzuki AM, Fletcher AR, Plachetzki DC, FitzGerald PC, Artieri CG, Atallah J, Barmina O, Brown JB, Blankenburg KP, Clough E, Dasgupta A, Gubbala S, Han Y, Jayaseelan JC, Kalra D, Kim YA, Kovar CL, Lee SL, Li M, Malley JD, Malone JH, Mathew T, Mattiuzzo NR, Munidasa M, Muzny DM, Ongeri F, Perales L, Przytycka TM, Pu LL, Robinson G, Thornton RL, Saada N, Scherer SE, Smith HE, Vinson C, Warner CB, Worley KC, Wu YQ, Zou X, Cherbas P, Kellis M, Eisen MB, Piano F, Kionte K, Fitch DH, Sternberg PW, Cutter AD, Duff MO, Hoskins RA, Graveley BR, Gibbs RA, Bickel PJ, Kopp A, Carninci P, Celniker SE, Oliver B, Richards S. Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res.* 2014 Jul;24(7):1209-23. doi: 10.1101/gr.159384.113. PMID: 24985915; PMCID: PMC4079975. PubMed ID: [24985915](#)

Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, Hafen E, Leivers SJ, Partridge L. 2001 Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein FBrf0135946; PMID: 11292874 PubMed ID: [11292874](#)

Drosophila 12 Genomes Consortium, Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., ..., MacCallum, I. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450(7167): 203--218. PMID: 17994087 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 FBrf0129795; PMID: 10942616 PubMed ID: [10942616](#)

Goberdhan DC, Paricio N, Goodman EC, Mlodzik M, Wilson C. 1999. *Drosophila* tumor suppressor PTEN controls cell size and number by antagonizing the Chico/PI3-kinase signaling pathway. *Genes Dev.* 13(24):3244-3258. PMID: 10617573 PubMed ID: [10617573](#)

Gonzalez, J. N., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., Raney, B. J., Powell, C. C., Nassar, L. R., Maulding, N. D., Lee, C. M., Lee, B. T., Hinrichs, A. S., Fyfe, A. C., Fernandes, J. D., Diekhans, M., Clawson, H., Casper, J., Benet-Pages, A., Barber, G. P., Haussler, D., Kuhn, R. M., Haeussler, M., Kent, W. J. (2021). The UCSC Genome Browser database: 2021 update. *Nucleic Acids research.* 49(1), 1046-1057. PMID: 33221922 PubMed ID: [33221922](#)

Grewal SS. 2009 Insulin/TOR signaling in growth and homeostasis: a view from the fly world FBrf0207364; PMID: 18992839 PubMed ID: [18992839](#)

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 2013 Aug;8(8):1494-512. doi: 10.1038/nprot.2013.084. Epub 2013 Jul 11. PMCID: PMC3875132; PMID: 23845962 PubMed ID: [23845962](#)

Hietakangas V, Cohen SM. 2009 Regulation of tissue growth through nutrient sensing. PMID: 19694515 PubMed ID: [19694515](#)

Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* 2016 May 19;44(9):e89. doi: 10.1093/nar/gkw092. Epub 2016 Feb 17. PMCID: PMC4872089; PMID: 26893356 PubMed ID: [26893356](#)

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., Haussler, D. (2002). The Human Genome Browser at UCSC. *Genome Res.* 12, 996-1006. PMID: 12045153 PubMed ID: [12045153](#)

Kim, M.J., O'Connor, M.B. 2021 *Drosophila* Activin signaling promotes muscle growth through InR/TORC1-dependent and -independent processes. FBrf0247766; PMID: 33234715 PubMed ID: [33234715](#)

Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati PV, Goodman JL, Gramates LS, Millburn G, Strelets VB, Tabone CJ, and Thurmond J and the FlyBase Consortium (2021). FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res.* 49(D1) D899–D907. PMID: 33219682 PubMed ID: [33219682](#)

Millington, J.W., Brownrigg, G.P., Basner-Collins, P.J., Sun, Z., Rideout, E.J. 2021 Genetic manipulation of insulin/insulin-like growth factor signaling pathway activity has sex-biased effects on *Drosophila* body size. FBrf0248704; PMID: 33793746 PubMed ID: [33793746](#)

Navarro Gonzalez, J., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., Raney, B. J., *et al.* (2021). The UCSC Genome Browser database: 2021 update. *Nucleic Acids Research.* 49(1), 1046-1057. PMID: 33221922 PubMed ID: [33221922](#)

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

11/15/2022 - Open Access

Poltlove, R. M., Jacobs, A. R., Haft, C. R., Xu, P., Taylor, S. I., 2000. Characterization of *Drosophila* insulin receptor substrate. *The Journal of biological chemistry* 275(30), 23346–23354. PMID: 1080187 PubMed ID: [1080187](#)

Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, Nguyen N, Paten B, Zweig AS, Karolchik D, Kent WJ. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics*. 2014 Apr 1;30(7):1003-5. PMID: 24227676 PubMed ID: [24227676](#)

Rele CP, Sandlin KM, Leung W, Reed LK. (2020). Manual Annotation of Genes within *Drosophila* Species: the Genomics Education Partnership protocol. *bioRxiv* 2020.12.10.420521 doi: 10.1101/2020.12.12.420521 PubMed ID: [null](#)

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

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: Hannah Congleton: formal analysis, validation, writing - original draft, writing - review editing. Cole A. Kiser: formal analysis, validation, writing - original draft, writing - review editing. Patricia A. Colom Diaz: formal analysis, writing - review editing. Elizabeth Schlichting: formal analysis, writing - review editing. Dorothy A. Walton: formal analysis, writing - review editing. Lindsey J. Long: supervision, writing - review editing. Laura K. Reed: supervision, writing - review editing. Juan Carlos Martinez-Cruzado: supervision, writing - review editing. Chinmay P. Rele: data curation, formal analysis, methodology, project, software, supervision, validation, visualization, writing - review editing.

Reviewed By: GEP Review Panel

History: Received June 30, 2021 Accepted November 10, 2022 Published Online November 15, 2022 Indexed November 29, 2022

Copyright: © 2022 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: Congleton, H; Kiser, CA; Colom Diaz, PA; Schlichting, E; Walton, DA; Long, LJ; et al.; Rele, CP (2022). *Drosophila mojavensis – chico*. microPublication Biology. [10.17912/micropub.biology.000677](https://doi.org/10.17912/micropub.biology.000677)