



OPEN

DATA DESCRIPTOR

Multifaceted and extensive behavioral trajectories of genomically diverse *Drosophila* lines

Daiki X. Sato^{1,2}✉, Takahira Okuyama³ & Yuma Takahashi²✉

Detailed tracking data is essential to understanding the intricate mechanisms behind animal behavior. Here, we present a comprehensive dataset containing behavioral movies and trajectories from over 30,000 *Drosophila melanogaster* individuals across 105 genetically distinct strains, including 104 wild-type strains from the *Drosophila* Genetic Reference Panel, along with one visually impaired mutant. These data, categorized by genetic background, sex, and social context (isolated or in groups), were collected during 15-minute sessions that included five minutes of repeated looming stimuli to elicit fear responses. Additionally, our experimental design incorporated group experiments with randomly combined pairs of strains to investigate synergistic effects of group members on behavioral dynamics. Beyond enabling detailed analyses of genetic factors underlying locomotion, fear responses, and social interactions, this dataset provides a unique opportunity to examine individual behavioral variability within genetically identical flies. By capturing a broad spectrum of behaviors across different genetic and environmental contexts, these data serve as a valuable resource for advancing our understanding of how genetics, individuality, and group interactions shape animal behavior.

Background & Summary

Behavioral genetics lies at the intersection of understanding how inherited traits are expressed in complex behaviors, a foundational question in life sciences. The fruit fly, *Drosophila melanogaster*, is an ideal model for such a quest, as it can be easily cultured due to its short generation time and many offsprings and has a compact genome that contains many orthologous genes involved in human diseases. Additionally, a range of genetic tools enables us to manipulate the fly's nervous system and behavior. These features allow for high-throughput behavioral assays and omics studies, making it an exceptionally powerful model organism in the field^{1–4}. Building on this foundation, we present a behavioral dataset derived from 36,600 flies, encompassing 104 inbred, isogenic wild-type strains and a mutant strain. This extensive and unique dataset refines our understanding of how genetic variations influence behavior and integrates experimental techniques to simulate environmental threats, thereby enhancing the ecological validity of our observations.

At the core of this dataset is the methodological rigor of employing isogenic lines, ensuring that the observed behavioral traits accurately reflect genetic influences, rather than environmental or developmental variability. Specifically, we exposed flies to computer-controlled looming stimuli—simulating predatory threats—to capture a range of behaviors including baseline locomotor activity, exploration, freezing, and social interactions, quantified with advanced tracking software. Further employing the approach of genome-wide association studies (GWAS) enables us to dissect the genetic determinants of behavior, providing insights into how specific genetic variations affect collective behavioral phenotypes and enhancing our understanding of the genetic architecture that underlies complex behavioral traits.

Our research also includes experiments on mixed groups of genetically distinct strains, exploring the impact of genetic diversity on group behavior. This aspect of the study illuminates the role of interindividual genetic variation in shaping group-level behavioral dynamics, offering a unique perspective on how genetic diversity within

¹Institute for Advanced Academic Research, Chiba University, Chiba, Japan. ²Graduate School of Science, Chiba University, Chiba, Japan. ³Graduate School of Science and Engineering, Chiba University, Chiba, Japan. ✉e-mail: daiki.sato12@gmail.com; takahashi.yum@gmail.com

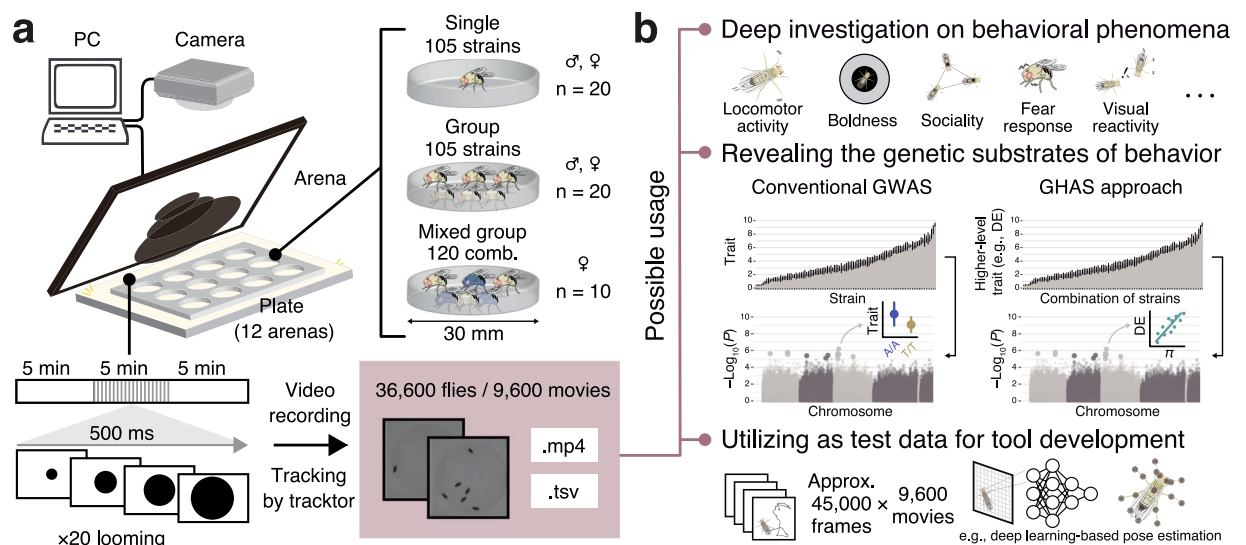


Fig. 1 Experimental workflow and example usages of the large-scale behavioral dataset of fruit flies. **(a)** The experimental setup in the current study. Utilizing over a hundred of strains and strain-pairs allows for detailed quantification of behavior. The provided dataset comprises of 9,600 files of raw movies and tracking data. **(b)** The usages of the dataset includes not only behavioral genetics research but also in-depth investigation of behavioral phenomena, as well as development of novel toolkits. Along with conventional GWAS approach using strain-level metrics, our mixed-strain approach allows for the quantification of higher-level traits [e.g., diversity effect (DE) arising from the genetic diversity within groups] and investigation on their genetic substrates with genome-wide higher-level association study (GHAS).

a group affects collective responses to environmental stimuli and underscores the evolutionary implications of genetic variability on social behavior.

By integrating sophisticated experimental designs with genetic uniformity and diversity, this dataset represents a significant advance in behavioral genetics (Fig. 1). It enhances our comprehension of the genetic basis of individual and collective behaviors in response to environmental challenges and sets the stage for future research into the interactions between behavior, genetics, neuroscience, and evolution. As a comprehensive resource, the dataset invites the scientific community to explore the genetic complexities of behavior, fostering innovative insights and methodologies in the quest to decipher the intricate genetic codes that govern animals' behavioral patterns.

Methods

Fly strains and keeping conditions. In our study, we utilized a selection of 104 strains from the *Drosophila* Genetic Reference Panel (DGRP), which reflects the natural genetic variation found within a population from North Carolina, US. To assess the role of visual information in our experiments, we included the visually impaired mutant strain *w[*] norpA[P12]*⁵. All strains were sourced from the Bloomington *Drosophila* Stock Center, with the specific strains and their stock numbers detailed in Supplementary Table 1. The flies were maintained at 25°C and 40–60% humidity, following a 12 L:12D light-dark cycle, and were fed standard food (Bloomington Formulation, Nutri-Fly™, #66–113, Genesee Scientific). We tested adult flies from both sexes, aged 2–4 days post-eclosion, in our behavioral assays, with 20 flies per sex and social setting for each strain.

In addition to the single-strain trials, we also conducted experiments with mixed-strain groups to explore how genetic and behavioral diversity influences group dynamics. We chose 15 DGRP strains alongside the mutant strain and created groups consisting of six female flies, with three individuals from each of two different strains, resulting in 120 unique pairings ($n = 10$ for each pairing).

Behavioral assay under looming stimuli. Behavioral tests were carried out within the Zeitgeber period of 0–12 in an incubator that maintained consistent temperature and humidity levels similar to those used for housing the flies. The setup for the experiments was aligned with the configurations shown in Fig. 1a and followed methodologies referenced in prior studies⁶. We chose six individuals to be grouped since the previous study demonstrated the number to be enough to show group effect in the assay⁷. In brief, flies, either individually or in groups of six, were anesthetized using CO₂ and positioned in a specially designed arena with inclined walls (30 mm in diameter and 2 mm deep; Supplementary Data 1), which prevented the flies from flying. To minimize sexual behaviors that might interfere with the response to visual stimuli and interactions among flies, we employed single-sex groups in the group experiments. After a 30-minute acclimation period, which included a 5-minute recording session, the flies were subjected to looming stimuli for 5 minutes. This involved 20 episodes of a black circle expanding over 500 ms on a white background, occurring every 15 seconds as previously described⁶. This was followed by another 5-minute unstimulated session, allowing for potential investigations into the recovery from the fear response. The stimuli were displayed at a 60 Hz refresh rate on a 13.3-inch EVICIV monitor,

which was angled at 45 degrees toward the experimental platform. This platform featured 12 separate arenas and was equipped with an LED board (147 mm wide \times 115 mm high, 7500 lx, TLB-MP, Asone Co., Japan) positioned underneath. Flies from each experimental group—differentiated by strain, sex, and social setting—were randomly assigned to one of the arenas. Behavioral responses were captured using a USB3 camera (DMK33UX290, The Imaging Source, Germany) at a resolution of 640 \times 480 pixels, recording at approximately 50 frames per second, and managed via a custom Python script. Example videos for the three (single, group, and mixed-group) conditions of flies are provided as Supplementary Videos 1–3.

Behavioral tracking and calculation of behavioral indices. Recorded videos were trimmed for each arena, and flies' behavior was primarily tracked by the Tracktor software⁸. However, we identified tracking errors that required corrections. To address these issues, we built an additional custom Python script. This script corrects: (1) instances where movement exceeded a threshold of 30 pixels per frame, which corresponds to roughly triple the fly's body length and suggests implausible movements rather than actual fly activity; (2) errors in identity swaps among individual flies due to crossing paths or overlapping trajectories; and (3) tracked points erroneously located outside the designated arena area, which was supposed to be masked to exclude from analysis. To rectify these issues, the script automatically reverted the coordinates of the affected flies to those of the previous frame whenever any of these conditions were met, ensuring the consistency and reliability of our behavioral data for subsequent analysis. To further reduce tracking calibration errors, we used the weighted moving average of coordinates computed over a time span consisting of the time point of interest and the preceding 8 time points (i.e., 9 time points in total). These coordinates were then used to calculate, for each time point, the distance traveled from the previous time point, moving speed (distance traveled divided by frame intervals), and head angle (determined from the displacement between time points 50 before and 50 after the time point of interest). The resulting measures were averaged at 0.5-s intervals for subsequent analyses.

Based on the coordinates at 0.5-second intervals, we calculated several time-averaged behavioral indices, including moving speed (traveled distance divided by frame intervals; mm/s), and nearest neighbor distance (NND; mm) for the first 5 minutes as indices of locomotor activity and sociality, respectively. We also calculated the time (s) spent in the center of the arena (i.e., within 7.5 mm from the center) during the first 5 minutes as the index of boldness. In the middle 5 minutes of the experiment, the average time taken to restore the moving speed to the baseline level (i.e., the mean moving speed during the first 5 minutes) after exposure to stimuli was used as an indicator of the fear response to threatening stimuli and is hereafter called freezing duration (s). The summary statistics, including the mean and standard deviation, for these metrics across the 105 strains are detailed in Supplementary Table 2.

Validation of tracking performance and behavioral metrics. To examine the performance of our tracking procedure, we compared it to that of FlyTracker⁹, a broadly used software to track flies' behavior with high accuracy. We used randomly chosen, independently experimented 100 movies of grouped flies as a validation dataset and tracked flies by tracktor with or without our correction script and counted frames with mismatches in flies' identity compared to the results of FlyTracker. This calculation was performed by comparing the coordinates of individuals tracked by each software, assuming that the closest individuals were the same entity. Mismatches were considered to have occurred when this correspondence changed over time. To further check the behavior-level consistency across software, we compared the above 4 types of time-averaged behavioral indices (i.e., moving speed, NND, time spent in the arena center, and freezing duration).

We also examined the correlation of locomotor activity, a fundamental behavioral metric, measured in the present study with publicly available activity metrics derived from previous studies. We analyzed five datasets as follows: (1) Total activity during time spent awake for 7 days¹⁰, referred to as “Harbison *et al.* 2013, waking activity.” (2) Total activity of untreated flies for 2 days¹¹, referred to as “Zhou *et al.* 2016, locomotor activity.” (3) Total distance traveled by control flies over 10 minutes¹², referred to as “Rohde *et al.* 2019, locomotor activity.” (4) Basal activity of untreated flies over one hour¹³, referred to as “Watanabe *et al.* 2020, basal activity.” (5) Climbing activity of untreated flies¹⁴, referred to as “Watanabe & Riddle 2021, climbing activity.” Datasets 1–3 and 5 were obtained from DGRP¹⁵. We used behavioral data from single-conditioned flies in the present study and examined correlations among studies separately for each sex.

Data Records

The data records referenced in this study are stored and accessible on Dryad with DOI: 10.5061/dryad.ttdz08m60¹⁶. The dataset comprises raw videos and tracking data, organized into subdirectories compressed for each strain, named as “Experiment_Type_Strain.tar.gz”. “Experiment” refers to the experimental design, labeled as either single-strain (“1_monostrain”) or mixed-strain (“2_mixedstrain”), and “Type” indicates whether the files is a recorded video (“video”) or a tracking dataset (“track”). Each behavioral video, formatted in mp4 and sized at 160 \times 160 pixels, corresponds to a specific arena. The tracking data are available in two formats: the raw output from the Tracktor software⁸ and the corrected outputs via our custom script, stored as csv and tsv files, respectively. The “Strain” refers to the experimented strain in single-strain setups and the combination used in mixed-strain experiments. A detailed description of the fields included in the tracking data is provided within the repository. We additionally provide supplementary files associated with the present article in the repository, including supplementary tables regarding the strain information and per-strain behavioral metrics, the STL file of the experimental arena, and example videos of recorded flies.

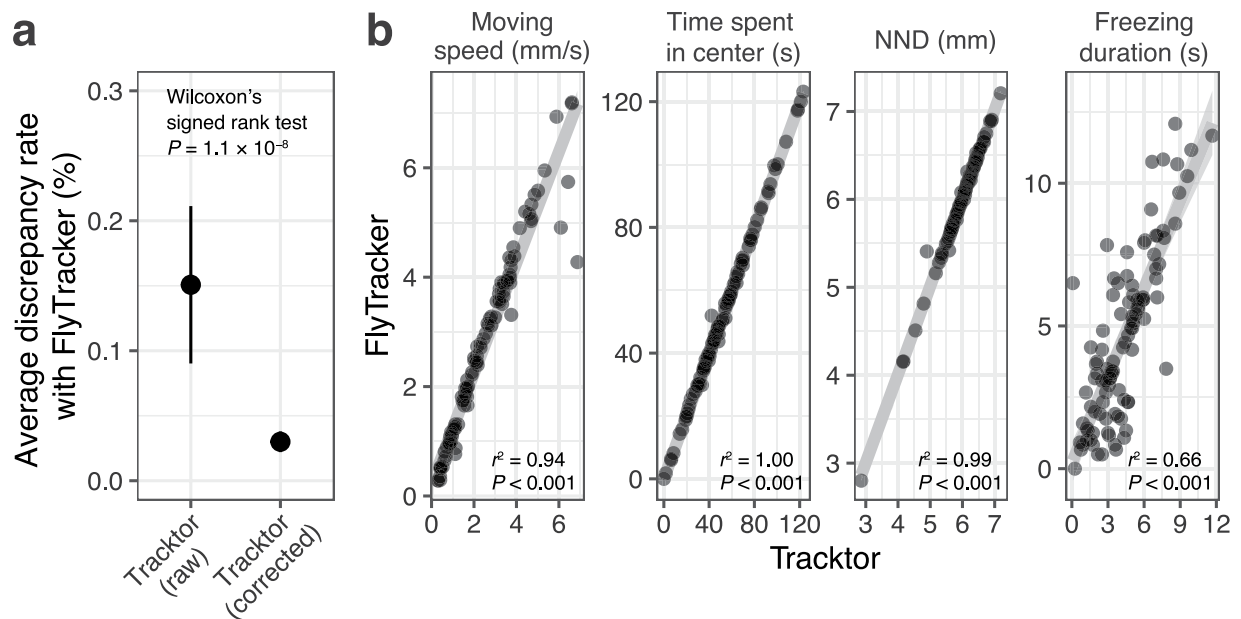


Fig. 2 Validation of trajectory data in comparison to that of FlyTracker. **(a)** The tracking data of randomly selected 100 videos compared between tracking systems, the number of mismatches in fly identity per frame was significantly lowered by our custom script. Statistical significance was evaluated by Wilcoxon's signed rank test ($n = 100$ for each). **(b)** Correlation in behavioral indices between the tracking data from two software.

Technical Validation

The technical validation of our dataset involved a rigorous initial screening to eliminate dead or injured individuals that could compromise data quality. Further, we manually double-checked recorded videos to minimize errors in behavioral tracking. For quantitative validation, we compared the tracking data obtained with FlyTracker⁹, a highly accurate software used for *Drosophila* behavior. Parallel tracking was performed on randomly selected grouped fly videos using both FlyTracker and Tracktor. We then assessed mismatches in fly identities per frame between the systems. Incorporating our custom correction script markedly enhanced the accuracy compared to raw Tracktor data ($P = 1.1 \times 10^{-8}$, Wilcoxon's signed rank test; Fig. 2a). Behavioral metrics such as average moving speed, nearest neighbor distance (NND), time spent in the center of the arena, and freezing duration were also compared across software. The results demonstrated a strong correlation for moving speed, time spent in the center, NND ($P < 0.001$, $R^2 = 0.94$, 1.00 , 0.99 , respectively; Fig. 2b). Although the correlation for freezing duration was lower, it remained adequately high ($P < 0.001$, $R^2 = 0.66$). Despite FlyTracker's high accuracy, its long processing times and occasional tracking failures led us to favor Tracktor. The precision achieved with Tracktor, augmented by our script, satisfactorily met our study's objectives.

We next investigated the correlation between locomotor activity for DGRP strains, measured as moving speed in the present study, and publicly available activity metrics obtained from previous studies^{10–14}. While the significant correlations were mostly observed in females (Fig. 3a), our metric showed a relatively high level of correlation with all previous studies (Fig. 3b,c), indicating that our dataset sufficiently captures behavioral patterns of DGRP lines. The stronger correlation in females is consistent with previous studies (Fig. 3b,c), suggesting that locomotor activity tends to be more stable and less variable in females compared to males. Males, in contrast, may exhibit context-dependent behaviors such as territoriality or courtship-like behaviors, which could introduce additional variability in behavior across different studies. These findings suggest that, depending on the type of behavior being investigated, females may provide a more suitable model for certain behavioral analyses, particularly those requiring stable and reproducible locomotor measures.

Usage Notes

The behavioral dataset provided in this study serves as a robust foundation for behavioral genetics research, including GWAS that aim to elucidate the genetic mechanisms underpinning a wide range of behaviors in *Drosophila*. The summarized behavioral indices for 104 DGRP strains are immediately applicable to the GWAS pipeline available on the DGRP2 website (<http://dgrp2.gnets.ncsu.edu>) or others. Utilizing the trajectory dataset, more complex metrics such as visual reactivity towards conspecifics and its potential genomic substrates can be estimated¹⁷. In addition to the conventional GWAS approach that uses strain-level metrics, our mixed-strain strategy enables what we call a genome-wide higher-level association study (GHAS). This approach seeks to identify associations between higher-level traits—for instance, the effect of genetic diversity on a group-level phenotype—and genetic metrics (e.g., nucleotide diversity π calculated among group members within a particular genomic window), ultimately pinpointing the loci potentially responsible for the observed group-level diversity effect (DE) (Fig. 1b). As an unprecedentedly large-scale behavioral dataset, it is ideally positioned for exploring a broad spectrum of behavioral phenomena and serves as a valuable resource for testing new toolkits developed for future research (Fig. 1b). In particular, this dataset facilitates innovative insights into

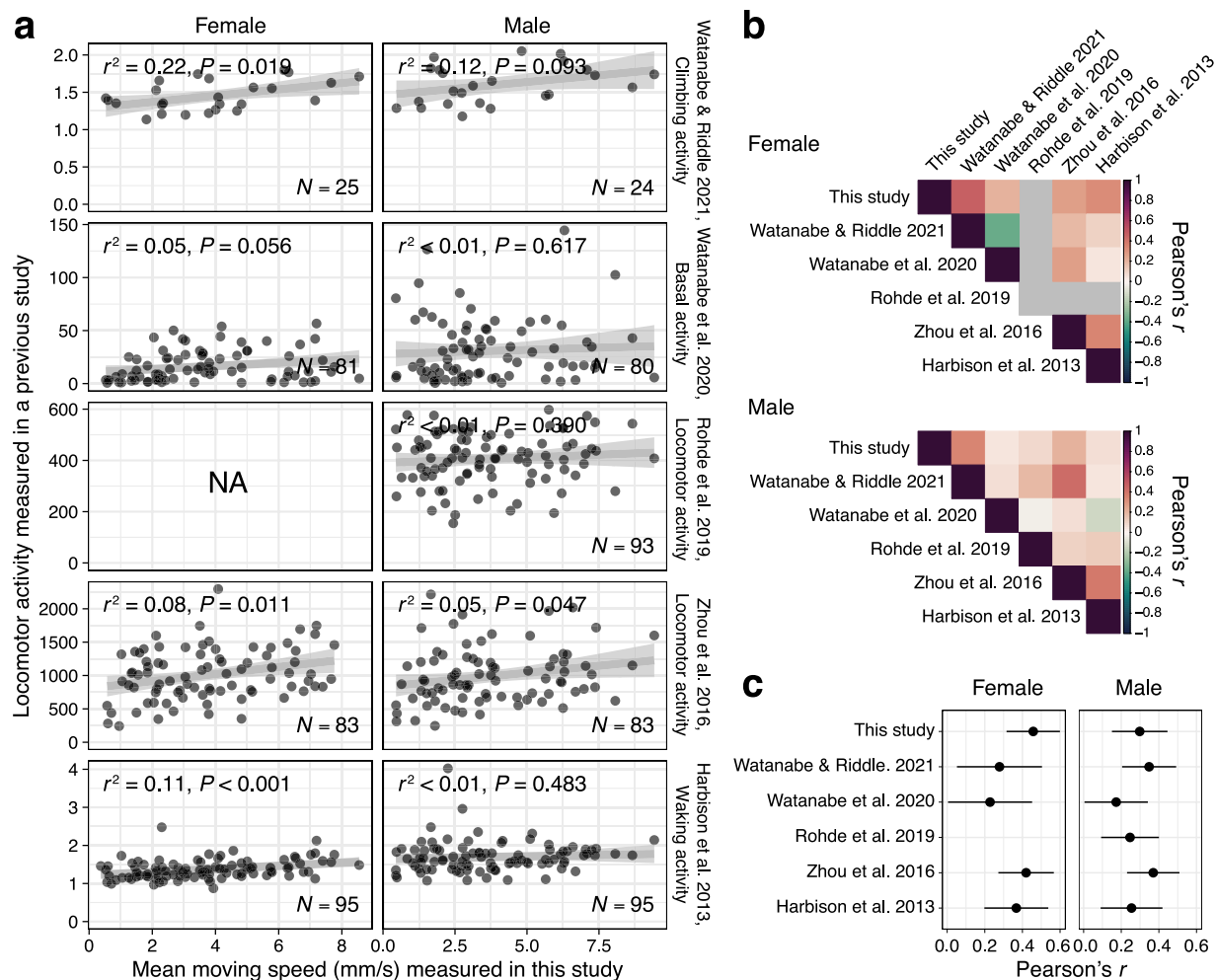


Fig. 3 Validation of locomotor activity in DGRP strains compared to previous studies. **(a)** The correlation between the mean moving speed of single flies measured in the present study and metrics related to locomotor activity reported in five previous studies^{10–14}. The dataset from Rohde *et al.*¹² includes only male data. The number of strains available for the calculation of correlation in each study pair are shown on the bottom of the panel. **(b)** Heatmap showing the correlations among the six studies. **(c)** Mean and standard error of correlations among studies.

animal behavioral dynamics by enabling comparisons between solitary and group conditions, as well as between sexes. Additionally, the assumption that individuals within a strain share identical genomic sequences provides a unique opportunity to examine behavioral variability arising from the same genetic background. This versatility allows researchers to tailor the dataset's use to their specific study objectives, making it a significant asset to the scientific community.

Code availability

The codes used in this study are available at https://github.com/daikisato12/Sato2025_Fruitfly_trajectory_data.

Received: 2 May 2024; Accepted: 27 February 2025;

Published online: 07 March 2025

References

1. Sokolowski, M. B. *Drosophila*: genetics meets behaviour. *Nat. Rev. Genet.* **2**, 879–890 (2001).
2. Ayroles, J. F. *et al.* Systems genetics of complex traits in *Drosophila melanogaster*. *Nat. Genet.* **41**, 299–307 (2009).
3. Branson, K., Robie, A., Bender, J., Perona, P. & Dickinson, M. High-throughput ethomics in large groups of *Drosophila*. *Nat. Methods* **6**, 451–457 (2009).
4. Geissmann, Q. *et al.* Ethoscopes: An open platform for high-throughput ethomics. *PLoS Biol.* **15**, 1–13 (2017).
5. Pak, W. L., Grossfield, J. & Arnold, K. S. Mutants of the visual pathway of *Drosophila melanogaster*. *Nature* **227**, 518–520 (1970).
6. Zacarias, R., Namiki, S., Card, G. M., Vasconcelos, M. L. & Moita, M. A. Speed dependent descending control of freezing behavior in *Drosophila melanogaster*. *Nat. Commun.* **9**, 3697 (2018).
7. Ferreira, C. H. & Moita, M. A. Behavioral and neuronal underpinnings of safety in numbers in fruit flies. *Nat. Commun.* **11**, 4182 (2020).
8. Sridhar, V. H., Roche, D. G. & Gingins, S. Tracktor: Image-based automated tracking of animal movement and behaviour. *Methods Ecol. Evol.* **10**, 815–820 (2019).

9. Eyjolfsson, E. *et al.* Detecting social actions of fruit flies. in *Computer Vision – ECCV 2014* 772–787 (Springer International Publishing, 2014).
10. Harbison, S. T., McCoy, L. J. & Mackay, T. F. C. Genome-wide association study of sleep in *Drosophila melanogaster*. *BMC Genomics* **14**, 281 (2013).
11. Zhou, S. *et al.* The genetic basis for variation in sensitivity to lead toxicity in *Drosophila melanogaster*. *Environ. Health Perspect.* **124**, 1062–1070 (2016).
12. Rohde, P. D. *et al.* Genetic signatures of drug response variability in *Drosophila melanogaster*. *Genetics* **213**, 633–650 (2019).
13. Watanabe, L. P., Gordon, C., Momeni, M. Y. & Riddle, N. C. Genetic networks underlying natural variation in basal and induced activity levels in *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics* **10**, 1247–1260 (2020).
14. Watanabe, L. P. & Riddle, N. C. Exercise-induced changes in climbing performance. *R. Soc. Open Sci.* **8**, 211275 (2021).
15. Gardeux, V. *et al.* DGRPpool: a web tool leveraging harmonized *Drosophila* genetic reference panel phenotyping data for the study of complex traits. *Elife* **12** (2023).
16. Sato, D. X., Okuyama, T. & Takahashi, Y. Data from: Multifaceted and extensive behavioral trajectories of genomically diverse *Drosophila* lines. <https://doi.org/10.5061/dryad.ttdz08m60> (2025).
17. Sato, D. X. & Takahashi, Y. Neurogenomic diversity enhances collective antipredator performance in *Drosophila*. *bioRxiv* 2024.03.14.584951, <https://doi.org/10.1101/2024.03.14.584951> (2024).

Acknowledgements

Computations for behavioral analysis were partially performed on the NIG supercomputer at ROIS National Institute of Genetics. This work was supported by the Japan Society for the Promotion of Science (Grants-in-Aid for Scientific Research JP22K15181 to D.X.S. and JP23H03840 to Y.T.), the Japan Science and Technology Agency (Strategic Basic Research Programs ACT-X JPMJAX24L7 to D.X.S.), and the Sasakawa Scientific Research Grant from The Japan Science Society (to D.X.S.).

Author contributions

D.X.S. and Y.T. conceived and designed the study. D.X.S. and T.O. collected and analyzed behavioral data of fruit flies. D.X.S. and Y.T. acquired the funding. D.X.S. wrote the first draft of the manuscript and all authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-04724-3>.

Correspondence and requests for materials should be addressed to D.X.S. or Y.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025