Reviewed Preprint
v1 • February 4, 2026
Not revised

Reviewed Preprint
v2 • August 5, 2026
Revised by authors

Simple Methods to Acutely Measure Multiple Timing Metrics among Sexual Repertoire of Male *Drosophila*

Yutong Song[1,†], Hongyu Miao[1,†], Dongyu Sun[1,†], Xiao Liu[1,†], Fan Jiang[2,†], Xuejiao Yang[1], Woo Jae Kim[1]

[1]HIT Center for Life Sciences, School of Life Science and Technology, Harbin Institute of Technology, Harbin, China • [2]University of Science and Technology of China, Hefei, China

For correspondence: wkim@hit.edu.cn

[†] These authors contributed equally.

Competing interests: No competing interests declared

Funding: See page 25

Reviewing editor: Mani Ramaswami, Trinity College Dublin, Ireland

eLife Assessment

This **useful** paper describes a software tool, "DrosoMating", which allows automated, high-throughput quantification of 6 common metrics of courtship and mating behaviors in Drosophila melanogaster. The validity of the tool is **convincingly** shown to perform as well as expert human assessments, across a wide range of conditions. While it is not suitable yet for posed based classification of individual actions, and its utility with respect to new deep learning-based tools such as DeepLabCut, SLEAP have yet to be assessed, the efficient coarse-grained detection of behavioral states in low resolution videos offered by DrosoMating represents a very helpful addition to analytical tools available for assessing courtship behaviour in Drosophila.

https://doi.org/10.7554/eLife.109742.2.sa3

Abstract

Male *Drosophila* courtship behavior is a key model for studying temporal decision-making. While courtship index (CI) is widely used to quantify mating activity, other timing metrics like courtship latency, copulation latency (CL), and mating duration (MD) remain understudied. Traditional methods for quantifying these behaviors are often labor-intensive and prone to human error. In this study, we present a protocol combining a modular chamber system and automated software (DrosoMating) to quantify 6 key timing metrics during male courtship. Our image-based video analysis enables precise identification of courtship and copulation events, as well as quantification of their timing and duration under controlled conditions. Validation shows <0.05% error rates and 98–99% agreement with manual scoring for CL, CI, and MD. The protocol supports genetic and neural circuit manipulations, detecting subtle genetic, social, and environmental behavioral variations. By minimizing manual effort and standardizing data collection, this approach facilitates scalable, reproducible studies on adaptive trade-offs, learning, and neural mechanisms in mating behavior. This method streamlines timing analysis in male courtship, offering reproducible metrics for behavioral genetics.

Highlight

A high-throughput software pipeline for automated temporal profiling of Drosophila melanogaster mating behavior after brief user-guided calibration.

An accompanying, open-hardware platform that can be assembled at minimal cost while maintaining experimental rigor.

The system attains near-manual accuracy and outputs Temporal Measurement Parameters data that are readily adaptable to—and quantifiable within—diverse behavioral paradigms.

Introduction

Time perception is critical for survival, enabling organisms to estimate intervals between events and execute adaptive behaviors. Male *Drosophila* courtship behavior, a genetically well-characterized model (Singh and Singh, 2016), provides a robust framework for studying temporal control in action sequences. A widely used metric to quantify mating activity is the 'Courtship Index (CI)', defined as the percentage of time a male engages in courtship behaviors (e.g., following, wing vibration, abdominal curling) toward a female during a standardized observation window. CI reflects both the intensity and persistence of courtship, serving as a sensitive readout for dissecting genetic, environmental, or experimental perturbations on mating behavior, as demonstrated in studies investigating genetic mutations, environmental perturbations, and neuronal manipulations (Benton, 2015; Greenspan and Ferveur, 2000a; Kim, 2009; Pavlou and Goodwin, 2013; Yamamoto and Koganezawa, 2013).

Despite extensive research on courtship displays (Chen et al., 2024a), other critical metrics remain poorly characterized, especially as their representative characteristics as timing behaviors. 'Courtship latency' (time until a male initiates courtship after detecting a female) may reflect motivational state or sensory processing deficits but is rarely quantified in terms of genetics (Eastwood and Burnet, 1977a). 'Copulation Latency (CL)' (time from courtship initiation to successful mating) and 'Mating Duration (MD)' (length of mating itself) are similarly underexplored, though both likely impact reproductive success. Prolonged MD enhances paternity assurance by maximizing sperm transfer and suppressing female remating, but it simultaneously increases predation risks and energy expenditure (Bretman et al., 2009). Conversely, reduced MD may decrease a male's paternity share due to incomplete sperm transfer or female resistance, yet it reduces exposure to ecological threats (e.g., predators, rivals) and conserves metabolic resources (Lee et al., 2023a). These traits may be influenced by cryptic factors such as male age, rival presence, or environmental stressors (e.g., temperature), yet their genetic and neural bases are unclear. While studies focus on stereotyped courtship actions, these "temporal metrics" offer insights into decision-making, persistence, and adaptive trade-offs—areas ripe for investigation (Huang et al., 2024; Kim et al., 2012a, 2013a, 2016, 2024; Lee et al., 2022; D. Sun et al., 2024; Y. Sun et al., 2024; Wong et al., 2019; T. Zhang et al., 2024b, 2024a; X. Zhang et al., 2024).

Numerous software tools have been developed to measure male mating behavior in *Drosophila*, yet many laboratories still rely on manual methods. Traditional approaches, such as the "courtship wheel", have been widely used to assess male courtship activity (Hall, 1994; Neckameyer and Bhatt, 2016; Vosshall, 2007). However, quantifying the full repertoire of mating behavior metrics has historically been labor-intensive (Ejima and Griffith, 2007; Sokolowski, 2001). Recent advances in automation have addressed these challenges. For instance, Dankert et al. (2009) introduced a machine vision system to quantify male-female interactions, though its application has been more focused on studying male-male aggression (Dankert et al., 2009a). With the rapid development of machine learning and image processing, tools such as 'Ctrax' (Branson et al., 2009), 'idTracker' (Pérez-Escudero et al., 2014), Cadabra (Stern et al., 2015), 'FlyTracker' and MATLAB combined with 'JAABA' (Barwell et al., 2021; Kabra et al., 2013) have enabled precise quantification of male-female mating behavior. Recently, Chen et al. (2024) introduced an image-processing-based system that addresses this gap by automatically recognizing multiple courtship elements with high accuracy (>97%) and robust identity tracking even during complete fly overlap(Chen et al., 2024b). Despite these advancements, recent efforts to simplify courtship conditioning protocols (Gil-Martí et al., 2023a) still lack precise quantification for temporal parameters of post-copulatory behavior. Consequently, there is a need for more streamlined protocols and software to accurately measure the timing aspects of male mating behavior in *Drosophila*.

Beyond *Drosophila*-specific tools, general behavioral video analysis paradigms have been developed to track the movement and behavior of various organisms, including fruit flies. These methods often focus on tracking the position of single or multiple flies (Iyengar et al., 2012; Qu

et al., 2022; Simon et al., 2011) or using artificial markers (Gal et al., 2020). Additionally, programs like 'ToxTrac' (Rodriguez et al., 2017), 'UMATracker' (Yamanaka and Takeuchi, 2018), 'FIMTrack' (Risse et al., 2017, 2014), and 'ilastik' (Berg et al., 2019) have been employed to analyze body posture, behavior, location, and movement trajectories in diverse species, including flies, bees, mice, and zebrafish. However, these tools often fall short in capturing the nuanced courtship behaviors of *Drosophila*, as motion data alone may not sufficiently represent complex interactions.

To address these limitations, we have developed a specialized software suite called 'DrosoMating' for analyzing and quantifying timing behaviors within male *Drosophila* courtship. By establishing 6 key indicators related to male courtship, we provide a robust framework for quantifying male mating behavior metrics under controlled conditions. This approach significantly reduces the time and errors associated with manual identification. Through iterative optimization, our system achieves an error rate of less than 0.05% compared to manual scoring. We believe this toolset offers a novel and efficient method for studying timing in male courtship behavior, providing valuable data and theoretical insights while streamlining the analysis process.

Results

Video-Processing Solutions for Precise *Drosophila* Identification in Multi-Chamber Setups

Accurate identification of *Drosophila* individuals is essential for behavioral quantification given their small size. DrosoMating addresses this through three integrated video-processing features that enhance recognition fidelity in multi-chamber environments.

First, an effective recognition region function isolates individual wells as discrete analysis zones (Fig. 1A bottom), eliminating interference from adjacent chambers and background artifacts. Second, an image learning function adapts to phenotypic variability across strains. Operators designate three sample flies (Fig. 1B), enabling the algorithm to internalize strain-specific morphological signatures for improved detection. Third, a manual threshold adjustment function permits post-recognition optimization. Users visually validate automated identifications and dynamically refine sensitivity parameters (Fig. 1C), correcting false positives/negatives. Collectively, these features ensure robust tracking accuracy across diverse experimental setups and genetic backgrounds.

Development and Validation of DrosoMating for Automated *Drosophila* Mating Behavior Analysis

To address the labor-intensive nature, high error rates, and limitations in quantifying complex behavioral metrics during manual observation of *Drosophila* mating, we developed the software DrosoMating. This tool automates analysis of mating videos recorded in a customized chamber. DrosoMating processes videos to generate a CSV file containing 6 key behavioral metrics: courtship duration, MD, CI, courtship start time, mating start time, and mating end time. While these metrics are individually straightforward, they can be combined to quantify diverse aspects of mating behavior metrics.

DrosoMating enables quantification of specific mating behaviors based on research needs. We focused on three core indices:

1. **Courtship Index (CI):** Indicates male courtship investment per unit time. Higher CI values (reflecting longer durations of behaviors like wing vibration or female chasing) suggest stronger male courtship motivation or higher female attractiveness (Greenspan and Ferveur, 2000b).
2. **Copulation Latency (CL):** Measures the time from courtship initiation to copulation start. Shorter CL indicates efficient male courtship or high female receptivity. Longer CL may reflect male courtship deficits (Eastwood and Burnet, 1977a), intense female rejection, or environmental stressors (Greenspan and Ferveur, 2000b; Singh and Singh, 2014).

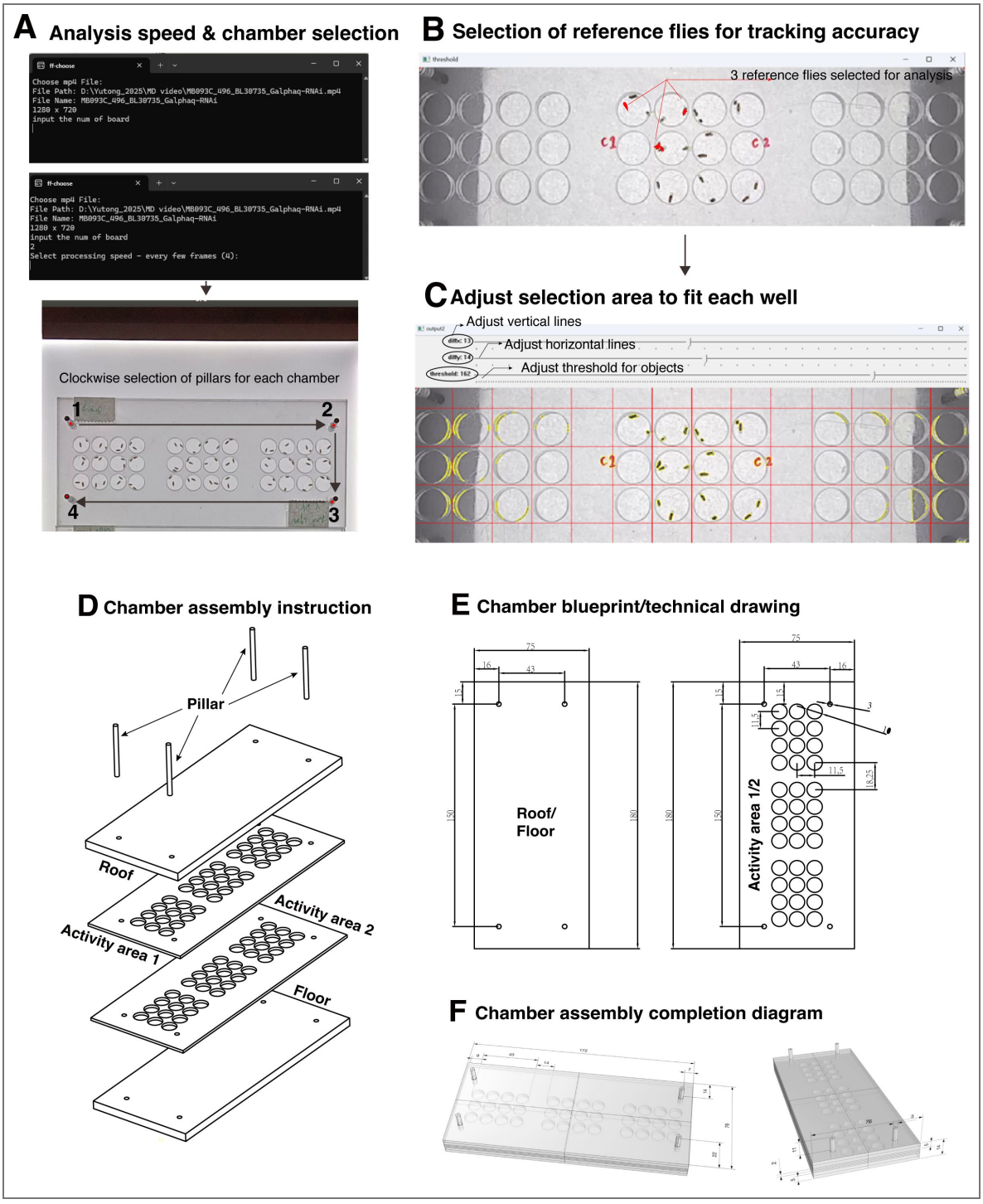

Figure 1. **Usage of DrosoMating and CAD Parameters of the Chamber**

(A) Typing input speed and input board numbers before selecting analyzing area (upper). Schematic representation of the region selection process (lower). Columns should be selected in a clockwise direction starting from the top left corner to define the chamber boundaries for perspective transformation (Fig. 1A, lower). Well numbering (1–36) follows a left-to-right, top-to-bottom sequence after perspective correction and is independent of pillar selection order. **(B)** Selection of reference flies for tracking accuracy. Any three reference flies were chosen for the analysis. **(C)** Use the X-axis, Y-axis, and threshold regulator to adjust the selection area to fit each well. Click "Enter" to run the program. Output file will be in the video folder: [Video Name]_output_[board numbers].csv (**D**) CAD (computer-aided design) exploded 3D view of the chamber instruction. (**E**) CAD parametric data of the chamber (dimensions in mm). (**F**) 3D diagram of the completed chamber

3. **Mating Duration (MD):** Represents the copulation period length. Longer MD may enhance sperm competition but increases predation risk and energy costs, illustrating an evolutionary trade-off between reproduction and survival (Greenspan and Ferveur, 2000b).

Together, CI, CL, and MD provide a robust framework for characterizing *Drosophila* behavioral states during mating. Accurate quantification of these metrics is essential.

Because overlapping behaviors occur frequently during courtship and mating, DrosoMating handles short and prolonged overlaps differently. When flies overlap for ≤15 frames, centroid velocity continuity is used to assign identity; if overlap >15 frames, the region is flagged 'occluded' and excluded from CI calculation. Occluded frames are excluded from CI calculation but retained for mating-state detection via merged-contour analysis, ensuring continuous MD measurement through copulation.

We validated DrosoMating by comparing its output for CI, CL, and MD against values from repeated, corrected manual measurements. Differences between software-generated data and manual data were consistently within 10 seconds for all three metrics, with no statistically significant differences (Fig. 2A-C). Manual quantification of these metrics requires several hours of effort and is prone to substantial human error. In contrast, DrosoMating automates the analysis. After a few minutes of initial setup and configuration, the software generates precise behavioral metrics from mating videos (Fig. S3).

Validation of DrosoMating Across Diverse *Drosophila* Strains and Behavioral Paradigms

DrosoMating was adapted to quantify mating behavior metrics across diverse *Drosophila* strains and experimental conditions. To evaluate its universality and accuracy, mating interactions were analyzed between virgin females and males of four strains: Canton-S (wild-type), Oregon-R (wild-type), w^{1118} (white eye mutant), and y^{1} (yellow body color mutant). These lines have been extensively employed as standard inbred strains in research contexts for a prolonged period. Wild-type strains (Canton-S and Oregon-R) exhibited significantly higher CI, CL, and MD compared to w^{1118} and y^{1} mutants (Fig. 3A–C). This quantitative assessment revealed markedly reduced activity in mutant males across all three behavioral metrics, consistent with known phenotypic limitations: w^{1118} males displayed restricted courtship due to visual impairment (Krstic et al., 2013), while y^{1} males demonstrated suboptimal mating performance, aligning with prior reports of courtship deficits (Drapeau et al., 2006).

Notably, reduced basal locomotor activity in w^{1118} and y^{1} mutants has been well documented in previous studies, independent of courtship behavior (Drapeau et al., 2006; Krstic et al., 2013). Consistent with these reports, our tracking data show that single-housed w^{1118} and y^{1} males exhibit lower average velocity than Canton-S and Oregon-R controls (Fig. S4A). These general locomotor differences are insufficient to fully explain the observed courtship and mating timing phenotypes, indicating that additional courtship_related processes contribute to the observed behavioral differences.

The software further validated established behavioral paradigms—Longer Mating Duration (LMD) (Kim et al., 2013a, 2012a) and Shorter Mating Duration (SMD) (Lee et al., 2023a)—using Canton-S males subjected to distinct pre-experimental conditioning. DrosoMating accurately detected expected differences in CL, CI, and MD between LMD (group vs. single housing) and SMD (sexually naive vs. sexually experienced) conditions (Fig. 3D–I), confirming its precision in resolving complex, paradigm-specific mating dynamics.

Quantification of Learning-Memory Dynamics and Neural Circuit Integration Using DrosoMating

DrosoMating facilitates the investigation of learning and memory consolidation in *Drosophila* through quantifiable behavioral metrics. The Learning Index (LI) was calculated as LI = (CI_naive − CI_trained) / CI_naive × 100%. (Fig. 4A), This metric provides a standardized measure of

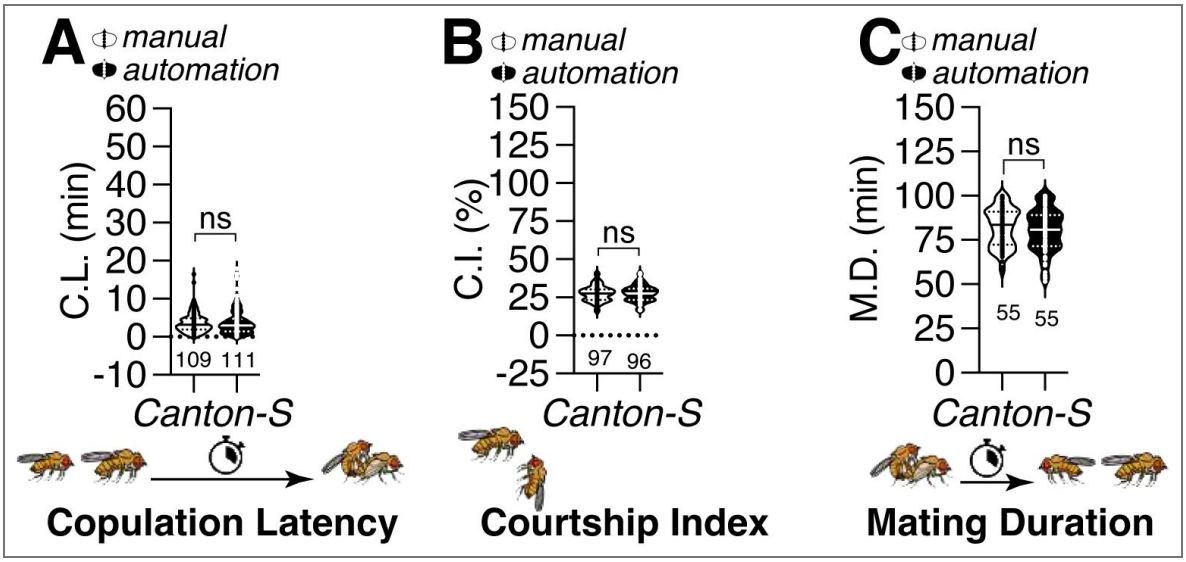

Figure 2. Comparison of the Accuracy of DrosoMating with Manual Scoring

(A) Copulation latency (CL) of Canton-S males was analyzed using DrosoMating and compared with manual recordings. For this figure, DBMs represent the 'difference between means' for the evaluation of estimation statistics. Asterisks represent significant differences, as revealed by the Student's t test (** p<0.05, ** p<0.01, *** p<0.001*). Differences between methods were assessed using two-sided Student's t-test (ns, not significant). See the Background for a detailed description of the CL assays used in this study. **(B)** Courtship index (CI) of Canton-S was analyzed by DrosoMating and manual recording. CI was measured by [courtship time/ (mating time-courtship time) *100%]. For detailed methods, see the BACKGROUND for a detailed description of the CI assays used in this study. **(C)** Mating Duration (MD) of Canton-S was analyzed by DrosoMating and manual recording. See the BACKGROUND for a detailed description of the MD assays used in this study.

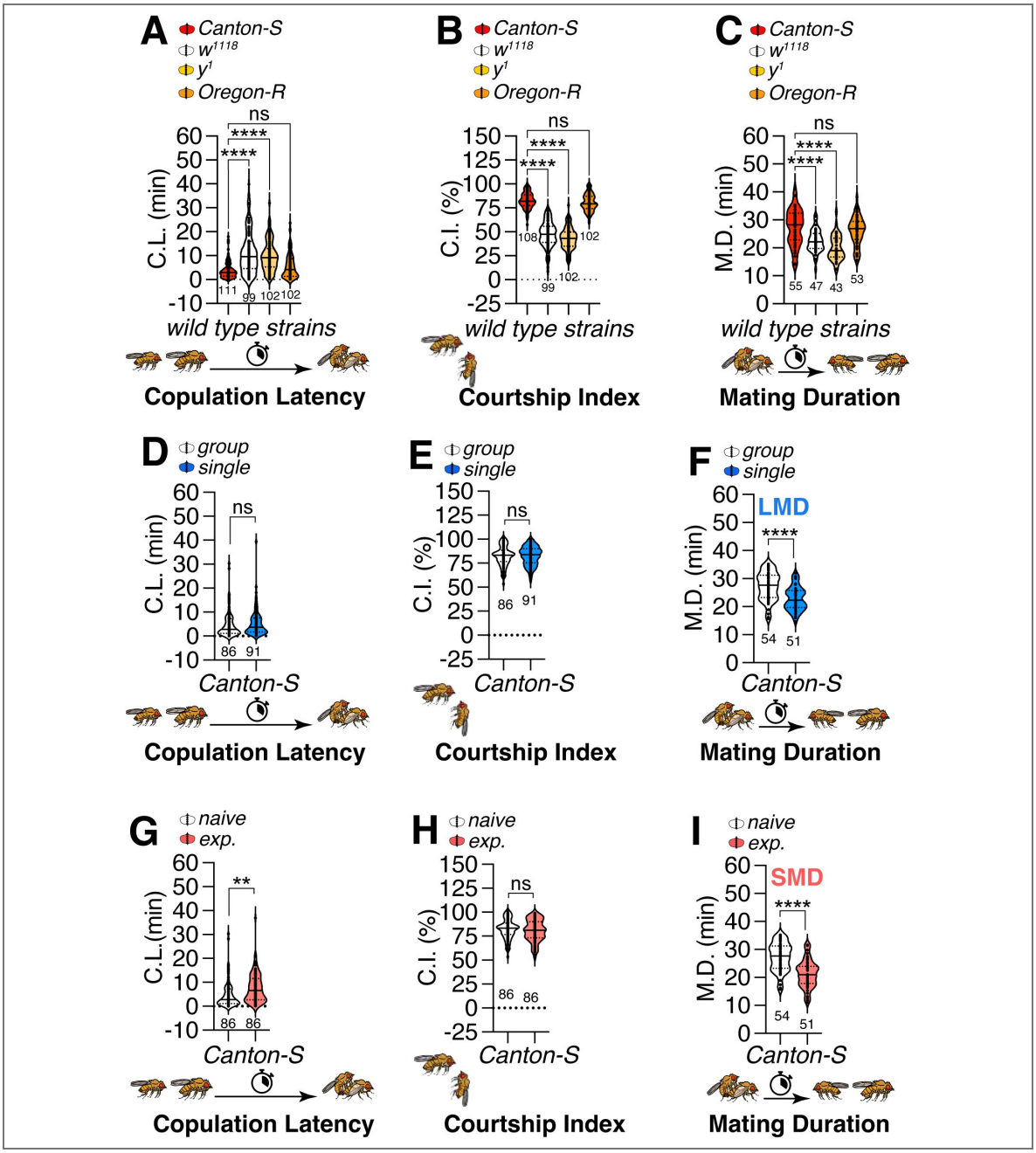

Figure 3. Behavioral assays of *Drosophila* males under different rearing conditions.

(A) CL of Canton-S, w^{1118}, y^{1}, and Oregon-R males. **(B)** CI of Canton-S, w^{1118}, y^{1}, and Oregon-R males. **(C)** MD of Canton-S, w^{1118}, y^{1}, and Oregon-R males. For detailed methods, see the background for a detailed description of the mating assays used in this study. Violin plots show the distribution of copulation latency (CL, in minutes) for Canton-S, w^{1118}, y^{1}, and Oregon-R males. Sample sizes are indicated below each group. Statistical significance was assessed using one-way ANOVA followed by Tukey's post hoc test for multiple comparisons for Fig. 3A-C. ****p < 0.0001; ns, not significant. **(D)** CL of group-housed and single-reared Canton-S males. For Fig. 3D-I, DBMs represent the 'difference between means' for the evaluation of estimation statistics. Asterisks represent significant differences, as revealed by the Student's t test (* *p<0.05,* ** *p<0.01,* *** *p<0.001*). Differences between methods were assessed using two-sided Student's t-test (ns, not significant). **(E)** CI of group-housed and single-reared Canton-S males. **(F)** LMD assays of Canton-S males. In the MD assays, white data points denote males that were group-reared (or sexually naïve), whereas blue data points signify males that were singly reared. The dot plots represent the MD of each male fly. The mean value and standard error are labeled within the dot plot. **(G)** CL of naive and sexually experienced Canton-S males. **(H)** CI of naive and sexually experienced Canton-S males. **(I)** SMD assays of Canton-S males. White data points represent sexually-naïve males and pink data points represent sexually-experienced ones.

experience-dependent behavioral modulation. This transformation of abstract behavior into discrete parameters enables rigorous analysis of neural and behavioral plasticity.

The software integrates seamlessly with neural manipulation techniques. For acute temporal control, thermogenetic tools (*UAS-shi^{ts}* (Kitamoto, 2001), *UAS-TrpA1* (Kang et al., 2012)) permit precise activation or silencing of circuits during courtship (Fig. 4B and S1A). Targeted perturbation of mushroom body Kenyon cells, critical for memory processing, revealed immediate behavioral consequences (Fig. S1B–C), elucidating their role in behavioral adaptation.

Established courtship conditioning paradigms (Gil-Martí et al., 2023a; Griffith and Ejima, 2009a; Kamyshev et al., 1999; Reza et al., 2013; Wolf et al., 1998) are robustly replicated, validating DrosoMating's fidelity in experience-driven behavior. (Fig. 4C). Mutant analyses revealed *orb2* (Fig. 4D–E) and *rut* (Fig. 4F-G) dependence of experience-induced CI reduction, consistent with known memory pathways.

Chronic neural interventions are enabled via *tub-GAL80^{ts}*-mediated adult-specific silencing or knockdown (Fig. S1D–F). This approach reveals enduring impacts of sustained circuit modulation on courtship plasticity and memory consolidation. Collectively, DrosoMating bridges behavioral quantification and mechanistic inquiry, advancing exploration of learning-memory interplay through precise, scalable experimental frameworks.

DrosoMating is more compatible with low-quality mating videos than conventional tracking-based pipelines

To evaluate whether conventional fly-tracking pipelines could be used as alternative tools for extracting mating-duration metrics, we tested Ctrax and FlyTracker on representative low-quality single-chamber mating videos recorded under our standard high-throughput conditions (Fig. 5). Ctrax is designed to estimate the position and orientation of multiple walking flies while maintaining individual identities over time (Branson et al., 2009) (*https://ctrax.sourceforge.net/*), whereas FlyTracker aims to track fly pose, including position, orientation, body size, wing and leg positions, and to generate trajectory- and feature-based outputs for downstream behavior analysis (Eyjolfsdottir et al., 2014). Because these programs were not readily compatible with our full high-throughput behavioral recording setup, we first cropped the original videos and tested single-chamber videos containing one male and one female fly.

Using Ctrax, we observed that target detection was highly variable even within single-chamber videos. In representative frames, Ctrax could occasionally identify the two flies correctly (Fig. 5A). However, imperfect segmentation was frequently observed. In some frames, parts of the fly body were detected as additional targets, resulting in over-segmentation and an apparent increase in the number of detected flies (Fig. 5B,C). Conversely, when the male and female were close to each other or physically overlapped, the two animals were sometimes detected as a single target (Fig. 5D). These examples indicate that Ctrax detection was sensitive to the low contrast and overlapping fly bodies present in our mating videos. We then adjusted the Ctrax detection threshold to improve segmentation quality. Although threshold optimization improved detection in some frames, abnormal detections remained evident across randomly sampled frames (Fig. 5E). For example, some frames still showed incorrect target numbers, and a severe segmentation failure was observed in the lower-right example, where the detected objects did not correspond to two clearly separable flies. Thus, even after parameter optimization, Ctrax did not consistently maintain the expected two-target detection state in single-chamber mating videos.

The instability of Ctrax detection was also reflected in the trajectory output. In the first 500 frames, the generated trajectories were fragmented into multiple colored track segments rather than two continuous trajectories corresponding to the male and female (Fig. 5F). In addition, some trajectories extended outside the chamber boundary, indicating tracking errors and identity instability. Consistently, frame-by-frame quantification of detected target number showed that the detected object count did not remain stable at the expected value of two flies per chamber (Fig.

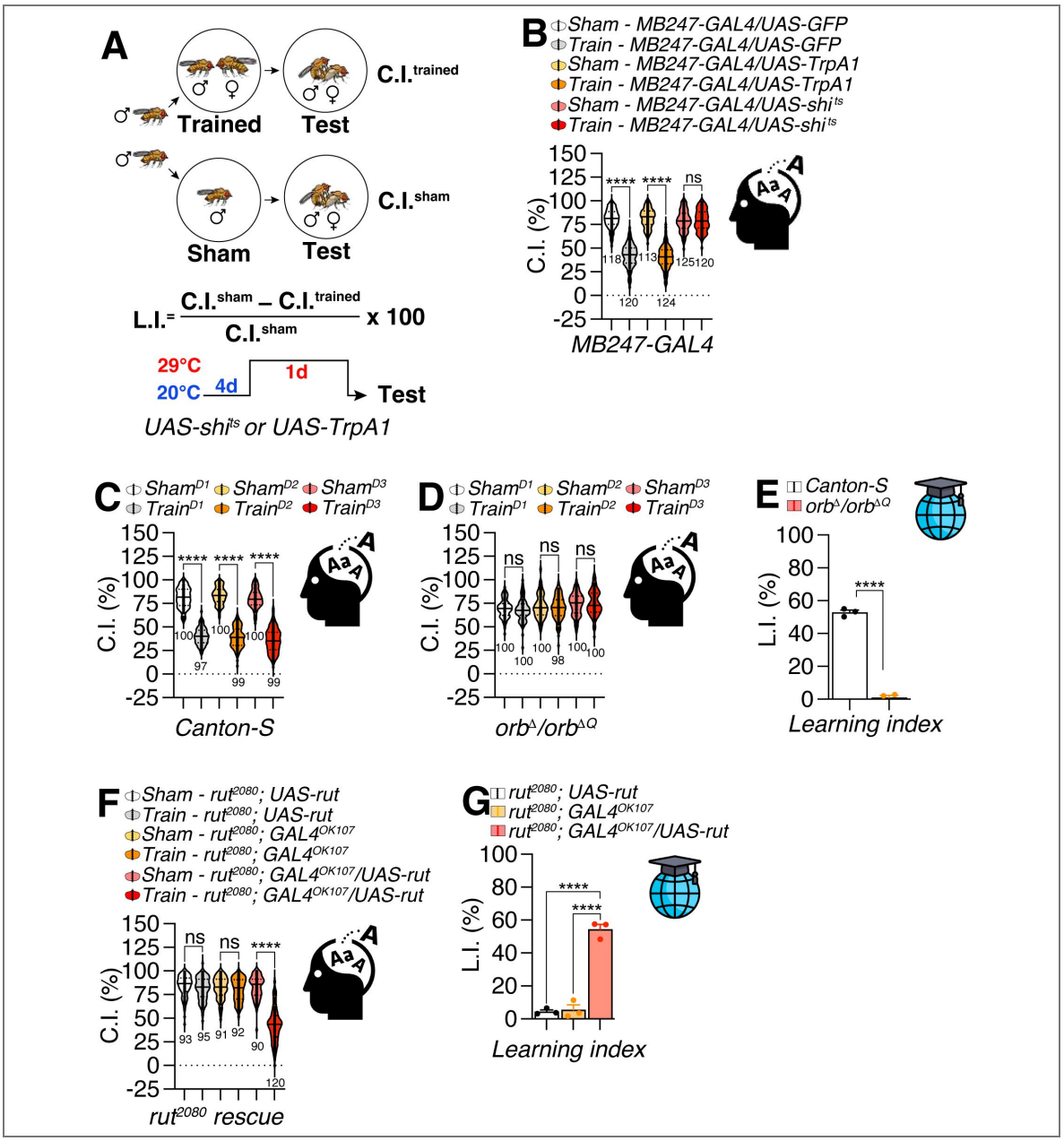

Figure 4. behavioral assays of learning in *Drosophila* which combined with temperature-dependent temporal activation/inhibition methods.

(A) Learning Index (LI) calculation and experimental design for temperature-dependent activation/inhibition of neural circuits in adult *Drosophila*. This figure illustrates the experimental paradigm for calculating the LI using CI measurements from sham and trained group. The protocol involves rearing flies at 20°C for four days followed by a one-day exposure to 29°C to activate/inhibit temperature-dependent, genetically-encoded neuronal modulators (for example, shi^{ts} as inhibitor or *TrpA1* as activator). **(B)** CI of training and sham training *MB247-GAL4/UAS-GFP*, *MB247-GAL4/UAS-TrpA1* and *MB247-GAL4/UAS-*shi^{ts} flies. Sample sizes (n) are indicated below each group. Statistical significance was assessed using two-way ANOVA for Fig. 4 B,C,D and F, followed by Sidak's post hoc test for pairwise comparisons between sham and trained groups within each genotype. ****p < 0.0001; ns, not significant. (**C**) CI of different training days (D1: 1 day, D2: 2 days, D3: 3 days) and sham trained Canton-S flies. (**D**) CI of different training days (D1: 1day, D2: 2 days, D3: 3 days) and sham training $orb^{\Delta}/orb^{\Delta Q}$ flies. (**E**) LI of Canton-S and $orb^{\Delta}/orb^{\Delta Q}$ males. For fig. 4E and G, DBMs represent the 'difference between means' for the evaluation of estimation statistics. Asterisks represent significant differences, as revealed by the Student's t test (* *p<0.05, ** p<0.01, *** p<0.001*). Differences between methods were assessed using two-sided Student's t-test (ns, not significant). (**F**) CI of training and sham training rut^{2080}*; UAS-rut,* rut^{2080}*;* $GAL4^{OK107}$*,* rut^{2080}*;* $GAL4^{OK107}$*/UAS-rut* flies. (**G**) LI of rut^{2080}*; UAS-rut,* rut^{2080}*;* $GAL4^{OK107}$*,* rut^{2080}*;* $GAL4^{OK107}$*/UAS-rut* flies.

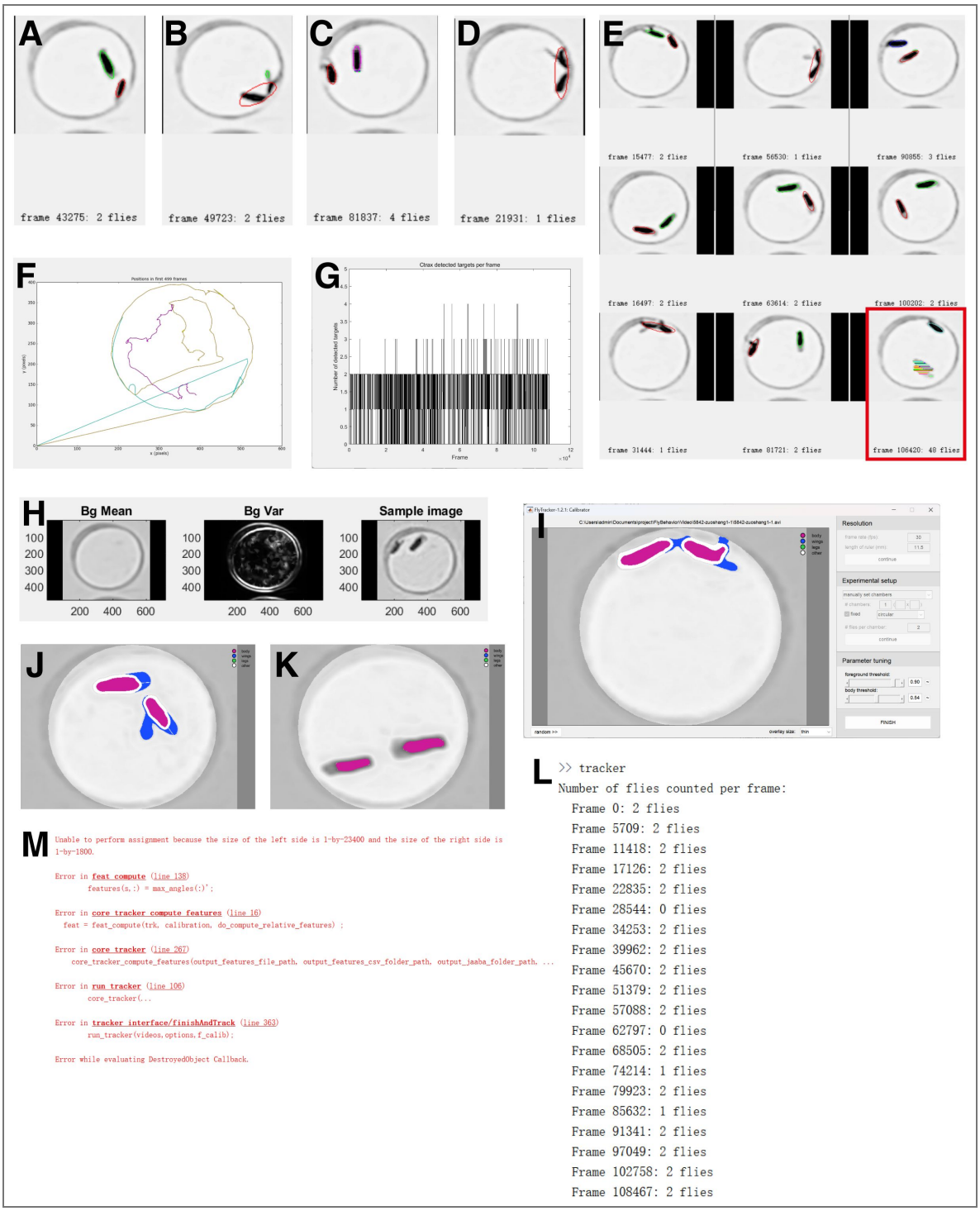

Figure 5. Limitations of conventional tracking-based tools for low-quality *Drosophila* mating videos.

(A-D) Representative Ctrax detection frames from single-chamber mating videos, showing examples with different detected target numbers. (E) Additional randomly sampled Ctrax detection frames during tracking. (F) Representative Ctrax trajectory plot generated from the tracked video. (G) Frame-by-frame plot of the number of targets detected by Ctrax. (H) FlyTracker background model, background variance, and sample image used for detection. (I-K) Representative FlyTracker segmentation outputs under different threshold settings. (I) Segmentation output using the default threshold setting. (J) Example frame showing relatively successful fly segmentation under the default threshold setting. (K) Representative segmentation output after lowering the threshold to reduce tracking/feature-computation errors. (L) FlyTracker command-window output showing the number of flies detected in sampled frames. (M) Example FlyTracker error message during feature computation.

5G). Because Ctrax failed to maintain stable two-fly detection and continuous trajectories under these video conditions, its output could not be reliably used for downstream extraction of copulation latency or mating duration.

We next tested FlyTracker on the same type of cropped single-chamber mating videos. FlyTracker was developed to track multiple flies by estimating body position, orientation, size, wing and leg positions, and by maintaining fly identities across video frames; it also outputs per-frame features such as velocity, facing angle, and wing-angle-related measurements for downstream behavioral analysis(Eyjolfsdottir et al., 2014). In our videos, FlyTracker was able to generate a background model, indicating that the program could recognize the overall imaging field and chamber background (Fig. 5H). However, during calibration and segmentation, the default threshold setting produced inconsistent detection results (Fig. 5I). Although some frames were segmented relatively well under the default threshold (Fig. 5J), these successful examples were not representative of the overall tracking process, and the program frequently terminated with runtime errors during tracking or downstream feature computation. To improve detection stability, we lowered the segmentation threshold. Under this adjusted setting, FlyTracker produced more complete fly masks in representative frames (Fig. 5K), and the diagnostic output showed that two flies were detected in many sampled frames (Fig. 5L). Nevertheless, detection remained unstable in some sampled frames, including frames in which zero or one fly was detected despite the expected two flies per chamber (Fig. 5L). The full pipeline ultimately failed during feature computation, producing a runtime error before complete tracking and feature outputs could be generated (Fig. 5M). Therefore, even after threshold adjustment, FlyTracker could not provide a stable end-to-end workflow for extracting mating-duration metrics from these low-quality mating videos.

This failure mode is relevant because FlyTracker depends on stable segmentation, identity maintenance, and per-frame feature extraction. In our assay videos, low contrast, chamber-edge artifacts, and prolonged male–female overlap during copulation interfered with these requirements. As a result, FlyTracker could occasionally identify the flies in individual frames, but it did not reliably complete the full analysis pipeline required for downstream behavioral quantification. This limitation is especially important for workflows such as JAABA, which use manually labeled examples to train behavior classifiers but still depend on upstream tracking-derived features. Thus, for our low-quality high-throughput mating recordings, FlyTracker-based analysis was substantially less practical than DrosoMating, which directly outputs mating-related timing metrics without requiring continuous high-fidelity two-fly pose tracking.

Methods and Materials

Fly Stocks and Husbandry

Drosophila stocks were maintained on standard cornmeal-agar food at 25°C and 60-70% relative humidity, with a 12-hour light/dark cycle. When suppression of temperature-sensitive transgene activity was required (e.g., for *UAS-shi^{ts}* or *UAS-TrpA1* experiments), flies were reared at 19-20°C throughout development. Because developmental duration at 19-20°C is approximately twice that at 25°C, experimental scheduling was adjusted to accommodate the delayed eclosion. Transgene activation was achieved by transferring flies to 29°C (*shi^{ts}*) or 28°C (*TrpA1*) for 30 mins prior to experiments. To eliminate confounding variables in behavioral assays, control and experimental groups were always raised under identical environmental conditions, including temperature, humidity, and light cycle.

Experimental models: Organisms/strains		
D. melanogaster: wild type: Canton-S	Bloomington Drosophila Stock Center	RRID: BDSC_64349
D. melanogaster: wild type: Oregon-R	Bloomington Drosophila Stock Center	RRID: BDSC_5
y^1 sc^1	Bloomington Drosophila Stock Center	RRID: BDSC_176
D. melanogaster: wild type: w^{1118}	Bloomington Drosophila Stock Center	RRID: BDSC_145

Df(1)Exel6234	Bloomington Drosophila Stock Center	RRID: BDSC_7708
orb2$^{\Delta}$	Bloomington Drosophila Stock Center	RRID: BDSC_94346
orb2$^{\Delta Q}$	Bloomington Drosophila Stock Center	RRID: BDSC_94347
UAS-TrpA1	Bloomington Drosophila Stock Center	RRID: BDSC_67132
tub-GAL80^{ts}	Bloomington Drosophila Stock Center	RRID: BDSC_7017
UAS-GFP.nls	Bloomington Drosophila Stock Center	RRID: BDSC_4775
UAS-shi^{ts}	Bloomington Drosophila Stock Center	RRID: BDSC_44222
UAS-KCNJ	Bloomington Drosophila Stock Center	RRID: BDSC_6596
UAS-TNT	Bloomington Drosophila Stock Center	RRID: BDSC_28838
UAS-NaChBac	Bloomington Drosophila Stock Center	RRID: BDSC_9469

GAL4^{OK107}	Bloomington Drosophila Stock Center	RRID: BDSC_854
rut^{2080}	Bloomington Drosophila Stock Center	RRID: BDSC_9404
UAS-rut *rut*^{2080}; *UAS-rut*	Bloomington Drosophila Stock Center Bloomington Drosophila Stock Center	Rw RRID: BDSC_9405
Y$^{hs\text{-}hid}$	Bloomington Drosophila Stock Center	RRID: BDSC_8846

Environmental Control System

The controlled-environment chamber is a dedicated fly room equipped to maintain standardized experimental conditions. It features a precision temperature control unit, which is a programmable apparatus for maintaining constant temperature with an accuracy of ±0.5°C. The chamber also includes automated humidity regulation through a humidifier or climate control system to stabilize relative humidity, typically within the range of 60–70%. Additionally, a programmable LED lighting system with full-spectrum arrays and an integrated timer is used to simulate circadian cycles, allowing for adjustable day/night duration and light intensity.

Mating Behavior Measurement System Setup

1. **Behavioral observation arena**: The behavioral observation arena was constructed from transparent acrylic plates. Each chamber measured approximately 1 × 1 × 0.2 cm unless otherwise specified. (Fig. 1D-E).
2. **Sex-separation films**: Non-permeable acrylic partitions (0.5 mm thickness) to isolate male and female cohorts prior to trials. Customized apparatus components can be fabricated by precision cutting of OHP (overhead projector) film sheets.
3. **Uniform illumination system**: Programmable LED panel array (full-spectrum, 6500K color temperature) positioned to optimize contrast for automated tracking software.
4. **High-resolution imaging**: OPPO A72 5G main camera (FHD/1080p resolution, 30 fps) mounted on an adjustable stabilizer for consistent video capture.

Mating Behavior Measurement Preparations

This protocol outlines the collection of male and virgin female *Drosophila melanogaster* and details the standardized mating behavior assay. To ensure experimental reproducibility, each cohort (genotype or condition) must include ≥100 individuals per sex, with final mating counts reaching at least 20 successful pairs (ideally 70-80) within a 2-hour observation window. The protocol spans developmental stages from egg to adult emergence, with adults requiring 5 days

post-eclosion to attain sexual maturity. Flies are reared under controlled environmental conditions (25°C, 60-70% relative humidity, 12-hour light/dark cycle) to minimize physiological variability. To avoid confounding effects of anesthesia on mating behavior, CO_2 exposure was strictly minimized during fly handling and sorting; brief, low-intensity exposure was reserved for initial virgin collection and before starting behavioral assays.

1. **Preparation of experimental male flies (15–20 days)**: Raise flies of the desired genotype (wild-type, mutants, or crosses such as GAL4/UAS combinations) in bottles at the specified temperature (standard: 25°C; alternative: 19°C or 29°C as required) under a 12:12 h light:dark cycle. Maintain 2–3 bottles per genotype, each containing 25 females and 15 males, to yield 25–30 flies per day. Transfer flies to fresh food every 2–3 days to prevent overcrowding.
2. **Post-eclosion handling**: Collect newly eclosed males (0–8 h post-eclosion) under brief CO_2 anesthesia. Isolate individual males in vials and age them for 5 days at 25°C (or 9 days at 18°C) to ensure complete brain maturation. Clear existing adults from culture vials daily using CO_2 to ensure synchronized collection of newly eclosed males. Group males into cohorts of 40 and rear in vials for 5–9 days to optimize sexual maturity while preserving mating efficiency.
3. **Large-scale virgin female collection for high-throughput genetic screening:** The *Df(1)Exel*^{6234}/*Y*$^{hs\text{-}hid}$ (Bloomington Stock Center #7708 with #8846) or the *SPR*^{attP}/*Y*$^{hs\text{-}hid}$ strain (Bloomington Stock Center #84576 with #8846) were used in this study. Both strains carry a null mutation in the *sex-peptide receptor (SPR)* gene (Yapici et al., 2008), which disrupts post-mating female refractoriness, thereby enhancing receptivity to remating and heat-inducible *hid* expression eliminating male progeny and balancer heterozygotes. Maintain cultures below 20°C prior to egg collection to prevent premature *hid*-induced lethality in *Y*$^{hs\text{-}hid}$ containing stocks. Higher temperatures (>20°C) may activate leaky *hs-hid* expression, reducing male lethality efficiency during subsequent heat shock.
   Place adult *Df (1)Exel*^{6234}/*Y*$^{hs\text{-}hid}$ flies in fresh bottles for 24 h to deposit eggs at 22°C. Heat shock incubate eggs at 37°C for 75 min to trigger *hid*-mediated apoptosis of males and balancer heterozygotes. We recommend using a water bath heated to 37°C for efficient heat shock instead of simply placing the samples in a 37°C incubator. Collect surviving female progeny and rear in groups of 40 for 5–8 days under standard conditions at 25°C to produce age-matched, virgin females. Raising at 29°C can speed up the process of obtaining virgin females. Following heat shock treatment, inspect all eclosed adults to confirm complete absence of male progeny. The *Df(1)Exel*^{6234}/*Y*$^{hs\text{-}hid}$ system should yield 100% female populations when properly executed.
4. **Chamber specifications:**
   - **Structure:** Reinforced with four stainless steel columns (3 mm diameter) to ensure stability (Fig. 1D-E and 6A).
   - **Material**: Four transparent acrylic plates (180 × 75 mm) with alternating thicknesses (1 cm and 0.3 cm), assembled in a "thick-thin-thin-thick" sequence (Fig. 1F and 6C).
   - **Layout**: 36 individual chambers arranged in 3×4 clusters (12 chambers per cluster).
   - **Chamber spacing**: 1 mm between individual chambers; 2 cm between clusters. Positioned 3 cm from the chamber's long edge and 1 cm from the short edge (Fig. 1E-F).
     Chamber dimensions are not fixed. By simply adjusting the magnification of the recording lens, users can re-scale the field of view and then refine the x (horizontal) and y (vertical) grid values in the DrosoMating interface to enclose the entire fly activity area within the corresponding wells. After this one-step calibration, the pipeline quantifies mating metrics (CL, CI, MD, etc.) with identical accuracy across different chamber sizes.
5. **Software implementation:** Automated analysis is performed using the DrosoMating pipeline (GitHub *https://github.com/hcls-kimlab/DrosoMating*) (Liu, 2024), the raw code is under win-ff module. which extracts courtship and mating metrics from video recordings. (Linux version: *https://github.com/liuxiao916/Fruit_Flies*)

Mating Behavior Assay Step-by-step Procedures

1. **Fly preparation**: On day 5 post-eclosion, select healthy males (of the desired genotype) and virgin females. Lightly anesthetize flies using ice or CO

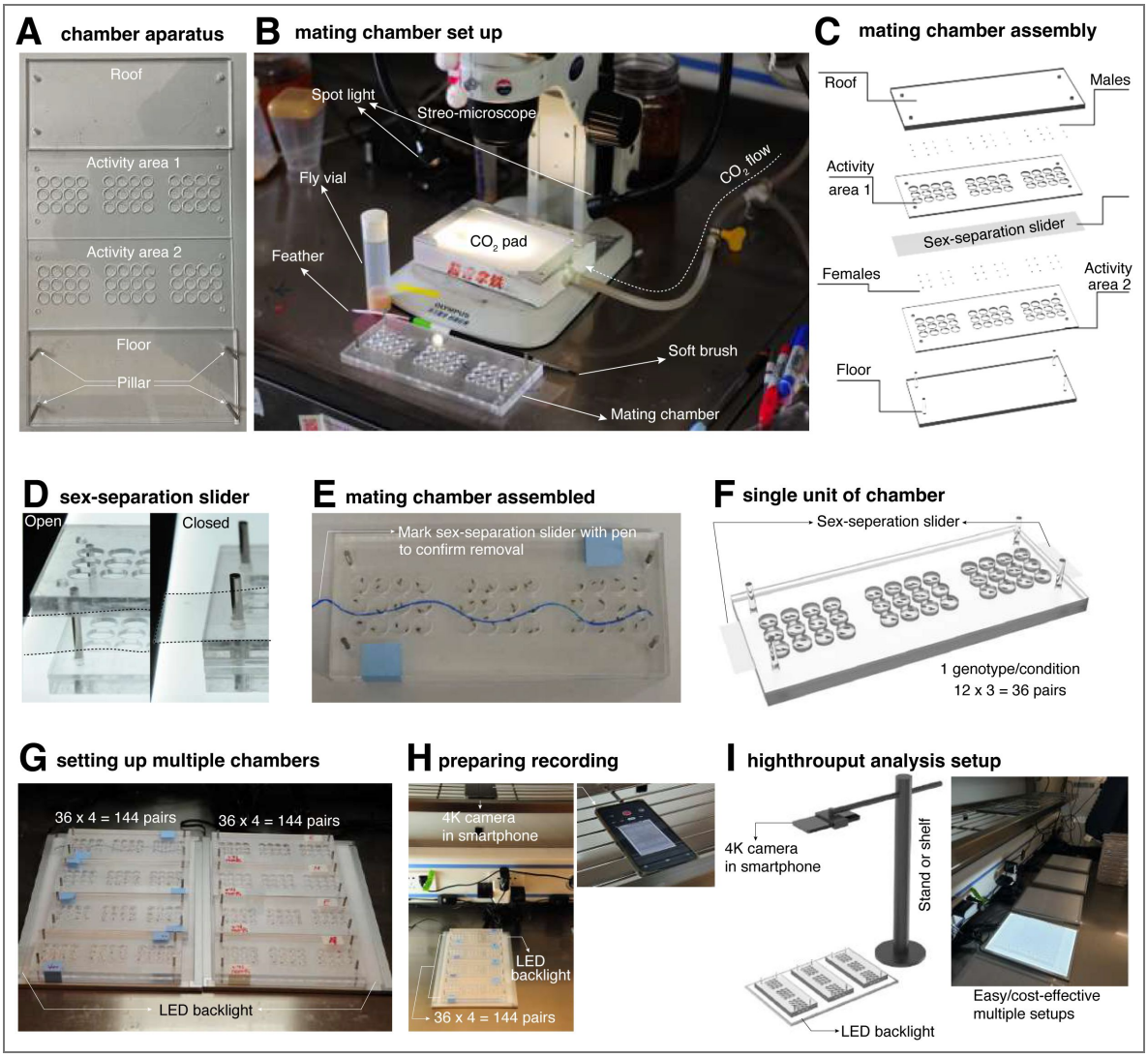

Figure. 6. Setup of mating assay chamber apparatus and architecture of workstation.

(**A**) Disassembly of the chamber apparatus. Each chamber comprises a roof, two activity areas, and a floor integrated with four pillars. (**B**) Workstation for assembling mating assay chambers. Essential required equipments for chamber assembly are shown: a stereomicroscope with an integrated light source, carbon dioxide anesthesia apparatus, a soft brush with feather on the opposite end, chamber apparatus, and fly containing vials. (**C**) 3D schematic of the mating chamber assembly. Assembly from bottom to top as shown. Briefly, first assemble the floor with one activity area 2 for placing females, then cover the females and separate activity area 1 from activity area 2 using the sex-separation slider. Next, position males and secure the chambers with paper tape. (**D**) The sex-separation slider is positioned between activity area 1 and activity area 2. (**E**) The mating chamber assembled with paper tape; the sex-separation slider should be marked with a pen to prevent forgetting to remove it. (**F**) 3D schematic of single chamber unit, can hold up to 36 pairs of fruit flies simultaneously. (**G**) Multiple chambers are set up with four chambers vertically aligned on an LED backlight, allowing the accommodation of 144 pairs of fruit flies across all chambers. (**H**) Video recording setup. Position a 1080p resolution camera or a smartphone with matching video capabilities above the chambers, turn on the LED backlight, and remove sex-separation slider before recording. (**I**) Setup for recording mating behavior using a camera mounted on a stand, positioned vertically above chambers arranged on a size-matched LED backlight to ensure uniform illumination. The camera is aligned to capture the entire chamber array, enabling simultaneous recording of mating interactions under controlled lighting conditions.

$_2$ pad (avoid prolonged exposure to prevent stress) (Fig. 6B). To minimize behavioral perturbations, flies should only be anesthetized at this step. For brief immobilization (e.g., sorting or transferring individuals), use CO_2 anesthesia (≤30 s exposure). To avoid bias, all experiments should be done blindly. Label fly tubes with random numbers instead of genotypes and keep a separate list matching numbers to genotypes. Only check which numbers correspond to which genotypes after finishing all data analysis.

2. **Chamber assembly and fly loading**: Using a soft brush, gently place a single female into a chamber hole in Floor (Activity Area 2) (Fig. 6B). Cover the chamber with a transparent film to prevent physical contact while allowing visual/chemical interactions. Align Activity Area 1 on top of the film (Fig. 6C). Introduce a single male into the opposite side of the film using the transparent film as a slider to minimize CO_2 exposure (Fig. 6D). Limit male exposure to CO_2 during this step to avoid behavioral artifacts. Seal the chamber securely with Roof using disposable paper tape (Fig. 6E).
3. **Recovery and Mating Initiation**: Transfer the assembled chamber to a 25°C incubator (humidity controlled at 60–70% RH) for 90 min to ensure flies recover fully from handling stress. After recovery, gently remove the transparent film to initiate physical mating (Fig. 6E and F).
4. **Video Recording Setup**: Position an adjustable LED panel beneath the chamber to ensure uniform illumination (Fig. 6G). Record mating behavior at 1080p resolution for 1.5–2 h (adjust duration as needed) (Fig. 6H). A minimal setup for capturing fruit fly mating behaviors can be rapidly established without extensive external facility requirements. The essential components include a camera stand, an LED light panel, and a camera capable of recording at a resolution and frame rate of 1080p (30 fps). (Fig. 6I).

Analysis of *Drosophila* Mating Video Data Analysis and Statistics

The use of the Windows version of DrosoMating

1. Prepare computer equipped either Windows (or Linux) PC with DrosoMating (ff-chose.exe) installed. The original code instructions for use are provided under the win-ff module.
2. Transfer video files into the computer to be analyzed. Launch DrosoMating and open the video file.
3. Enter the number of boards and velocity parameter. After entering the information, click “Enter” to proceed to the next step. (Fig. 1A)
4. Select the four steel marker pillars in clockwise order (Fig. 1A). Adjust the selection sequence based on the video orientation. For each plate, a preview image will appear. Click on three distinct flies to enhance tracking accuracy (Fig. 1B).
5. After selecting all regions, return to the DrosoMating home interface. Adjust x (horizontal) and y (vertical) values to align grids with chamber holes. Fine-tune fly detection by modifying the s value (sensitivity threshold) until all flies are correctly identified. Use the grid to segment individual flies (as shown in Fig. 1C).
   Note on the s value: The parameter s represents the grayscale intensity threshold (range: 0–255 for 8-bit images) used for binary segmentation of flies from the background. It is automatically calculated as the maximum grayscale value at the three reference fly positions plus an offset of 28, and can be manually adjusted to accommodate varying illumination conditions.
6. Run the analysis (Save the processed video but not recommended for routine use).
7. Export raw data as a CSV or Excel file. Open it in spreadsheet software for further processing.

The usage instructions for the Linux version of DrosoMating are presented in Fig. S2

Statistical Analysis

To ensure robust statistical analysis, each experimental group included at least 100 male flies (naïve, sexually experienced, or singly reared). Internal controls were incorporated into every experiment as recommended by Bretman et al. (2011) (Bretman et al., 2011). Normality of the mating duration data was confirmed using the Kolmogorov-Smirnov test (p>0.05). For group

comparisons, two-sided Student's *t*-tests were applied to calculate significance levels (****p<0.0001, ***p<0.001, **p<0.01, * p<0.05), comparisons among three or more groups were performed using one-way ANOVA with Tukey's HSD post-hoc tests. Two-factor experimental comparisons were performed using two-way ANOVA followed by Sidak's post-hoc test. Estimation statistics (Claridge-Chang and Assam, 2016) were additionally used to visualize effect sizes, mean differences, and precision, avoiding reliance solely on null hypothesis testing. All analyses, including data plotting, were performed using GraphPad Prism software.

Discussion

Advancing Courtship Behavior Analysis in *Drosophila*: From Manual Tracking to computer-vision classifier

The study of male *Drosophila* mating behavior represents a cornerstone in neurogenetics, providing critical insights into the neural and molecular mechanisms underlying innate social behaviors (Dauwalder, 2008; Dukas, 2020). Historically, the CI—quantifying the proportion of time spent in courtship activities—has served as a fundamental metric for evaluating mating activity (Greenspan and Ferveur, 2000a; Griffith and Ejima, 2009b; Pavlou and Goodwin, 2013; Sokolowski, 2001; Yamamoto and Koganezawa, 2013). However, traditional methodologies relied on labor-intensive manual observation, limiting their capacity to capture dynamic behavioral repertoires such as orientation, wing extension ("courtship song"), and copulation sequences (Dauwalder, 2008; Dukas, 2020). While machine vision systems like Dankert et al.'s automated platform advanced quantification of social interactions, their primary application skewed toward male-male aggression rather than nuanced courtship dynamics (Dankert et al., 2009b). Concurrently, research into genetic and environmental modulators—including *fruitless* (*fru*) and *doublesex* (*dsx*) gene pathways, cuticular hydrocarbons, and sensory cues—highlighted the need for scalable, high-resolution tools to dissect multimodal influences on behavior.

Technical Innovation and Validation

DrosoMating addresses these gaps through an integrated framework combining modular environmental chambers, automated video tracking, and machine-based annotation. This system minimizes manual intervention while capturing temporally precise metrics such as CL, CI, and MD. Validation against ground-truth manual scoring demonstrated exceptional accuracy (98–99% agreement; error rates <0.05%), underscoring reliability for high-throughput screens. Crucially, the platform detected subtle behavioral variations across genotypes—e.g., diminished courtship vigor in w^{1118} and y^{1} mutants—and environmental manipulations such as group-rearing or prior sexual experience, which respectively shortened mating duration by ~8–12 %. This sensitivity enables investigations into adaptive trade-offs, learning phenotypes, and circuit-level mechanisms. For instance, DrosoMating replicated classic deficits in aversive learning following mushroom body ablation and courtship impairments in *rutabaga* learning mutants, corroborating its utility for behavioral neuroscience (Levin et al., 1992).

Limitations and Future Directions

Despite its advantages, DrosoMating currently lacks granularity for posture-specific interactions (e.g., "wing threat" during aggression or "circling" during courtship) and multisensory integration. Future iterations could incorporate 3D pose estimation algorithms (e.g., DeepLabCut) or integrate JAABA-based classifiers to expand action repertoires (Kabra et al., 2013). Although DrosoMating does not currently export the full set of per-frame kinematic features required for direct JAABA classifier training, its modular video-processing framework could be extended in future versions to generate JAABA-compatible trajectory or feature outputs.

Cross-species testing was not performed in this study. Although DrosoMating's state-detection approach is morphology-agnostic and therefore theoretically applicable across *Drosophila* species, we have revised the text to avoid claiming demonstrated cross-species performance. Formal validation across diverse species remains a promising future direction.

Comparative Evaluation with Existing Courtship Analysis Tools

A major advantage of DrosoMating is its compatibility with low-quality, high-throughput mating videos. Conventional tracking-based tools such as Ctrax and FlyTracker are powerful for trajectory- and pose-based behavioral analysis, but they generally require stable object segmentation, identity maintenance, and reliable feature extraction across frames. Ctrax was designed to estimate the positions and orientations of multiple walking flies while maintaining their identities, whereas FlyTracker aims to track detailed fly pose and generate per-frame behavioral features. Under our recording conditions, these requirements were difficult to satisfy because the videos were low contrast and the male and female frequently overlapped during copulation.

In our tests, both tools showed limited compatibility with these videos. Even after cropping to single-chamber videos and adjusting detection parameters, Ctrax produced unstable target numbers, fragmented trajectories, and tracking errors. FlyTracker could generate a background model and occasionally segment flies successfully, but detection remained unstable and the full pipeline failed during feature computation. These issues are particularly relevant for copulation latency and mating duration analysis: during copulation, the male and female remain physically coupled for a long period, which makes identity-based tracking difficult. For these timing metrics, it is more important to robustly detect the onset and offset of the mating state than to reconstruct detailed individual trajectories. DrosoMating was purpose-built to address these specific challenges. It operates reliably on lower-quality video streams, requires no complex pre-processing or manual ROI definition, and is optimized for high-throughput multi-chamber analysis. While it does not offer the same level of pose or kinematic detail as other tools, it provides a unique solution for laboratories seeking a simple, fast, and robust pipeline to quantify core reproductive timing metrics—copulation latency (CL), courtship index (CI), and mating duration (MD)—without the overhead of more complex systems (Table.1).

Although we directly tested only Ctrax and FlyTracker, this limitation may also affect workflows that depend on upstream tracking-derived features. For example, JAABA uses tracking-derived features to train behavior classifiers, and DANCE, a recent *Drosophila* aggression and courtship pipeline, uses JAABA-based classifiers and lists FlyTracker and JAABA as required software. Thus, our conclusion is not that these tools are generally unsuitable for *Drosophila* behavior analysis, but that DrosoMating provides a more practical workflow for low-quality, high-throughput videos focused specifically on mating timing.

While DrosoMating currently prioritizes mating timing metrics over discrete behavioral classification, its modular architecture provides a foundation for future integration with behavior classifiers such as JAABA. Laboratories requiring granular behavioral elements—such as wing extension or circling—would benefit from an extended pipeline that exports per-frame kinematic features for downstream classifier training. Validating this integration represents a promising future direction to broaden the tool's utility beyond core reproductive timing assays.

	DrosoMating	Ctrax	FlyTracker
Performance on low-quality mating videos	Compatible	Unstable under tested conditions (require high contrast)	Unstable under tested conditions (require high contrast)
Multi-chamber analysis	Up to 4 chambers (36×4 wells)	Requires cropping or additional preprocessing	Requires cropping or additional preprocessing
Copulation overlap handling	Robust (state detection, no identity needed)	Unstable during overlap (over-/under-segmentation)	Unstable during overlap (Runtime error under tested videos)
Direct output of CL/CI/MD	Yes; automated CSV	No (unstable trajectories)	No (runtime error)
Accuracy vs. manual scoring	High agreement with manual scoring	Not quantified because no stable CL/MD output was obtained	Not quantified because no stable CL/MD output was obtained

Table 1. Comparison of DrosoMating with tracking-based tools under low-quality mating-video conditions.

The comparison was performed using representative low-quality, high-throughput mating videos recorded under standard experimental conditions. DrosoMating is compatible with low-contrast video streams, supports native multi-chamber analysis (up to 4 boards, 144 wells total), and robustly handles copulation overlap via state-detection without requiring individual identity maintenance. It directly outputs copulation latency (CL), courtship index (CI), and mating duration (MD) as automated CSV files with high (98–99%) agreement relative to manual scoring. By contrast, Under the tested low-quality mating-video conditions, Ctrax and FlyTracker did not provide stable end-to-end outputs suitable for reliable extraction of mating-timing metrics. DrosoMating was designed to extract mating-related timing metrics from high-throughput videos without requiring continuous two-fly identity tracking. Ctrax and FlyTracker were tested on cropped single-chamber videos from the same recording setup. Their limitations described here refer specifically to these low-quality mating videos and should not be interpreted as general limitations of the tools.

Supplementary figures

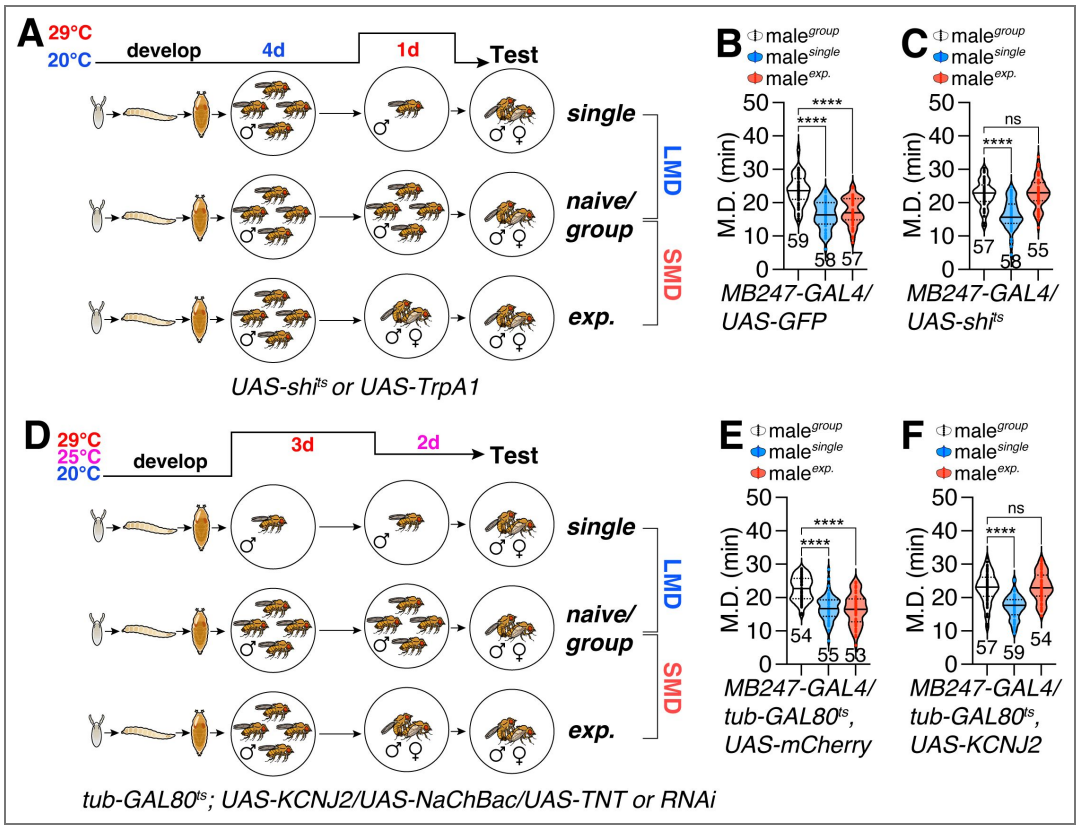

Figure S1. Temperature-dependent neural manipulation during LMD and SMD assays in adult *Drosophila*. (A) Schematic representation of LMD and SMD assays timeline when flies were crossed with heat-sensitive *Drosophila* cation channel *TrpA1* and *shi^{ts}*. **(B)** MD assays of flies expressing the *MB247-GAL4* driver together with *UAS-GFP*. Sample sizes are indicated below each group. Statistical significance was assessed using one-way ANOVA followed by Tukey's post hoc test for multiple comparisons for Fig. S1B-C,E-F. ****p < 0.0001; ns, not significant. **(C)** MD assays of flies expressing the *MB247-GAL4* driver together with *UAS-shi^{ts}*. **(D)** Schematic representation of LMD and SMD assays timeline when *tub-GAL80^{ts}; UAS-KCNJ2/UAS-NaChBac/UAS-TNT* or *RNAi* flies are crossed with specific GAL4 driver. **(E)** MD assays of flies expressing the *MB247-GAL4* driver together with *tub-GAL80^{ts}*, *UAS-mCherry*. **(F)** MD assays of flies expressing the *MB247-GAL4* driver together with *tub-GAL80^{ts}*, *UAS-KCNJ2*.

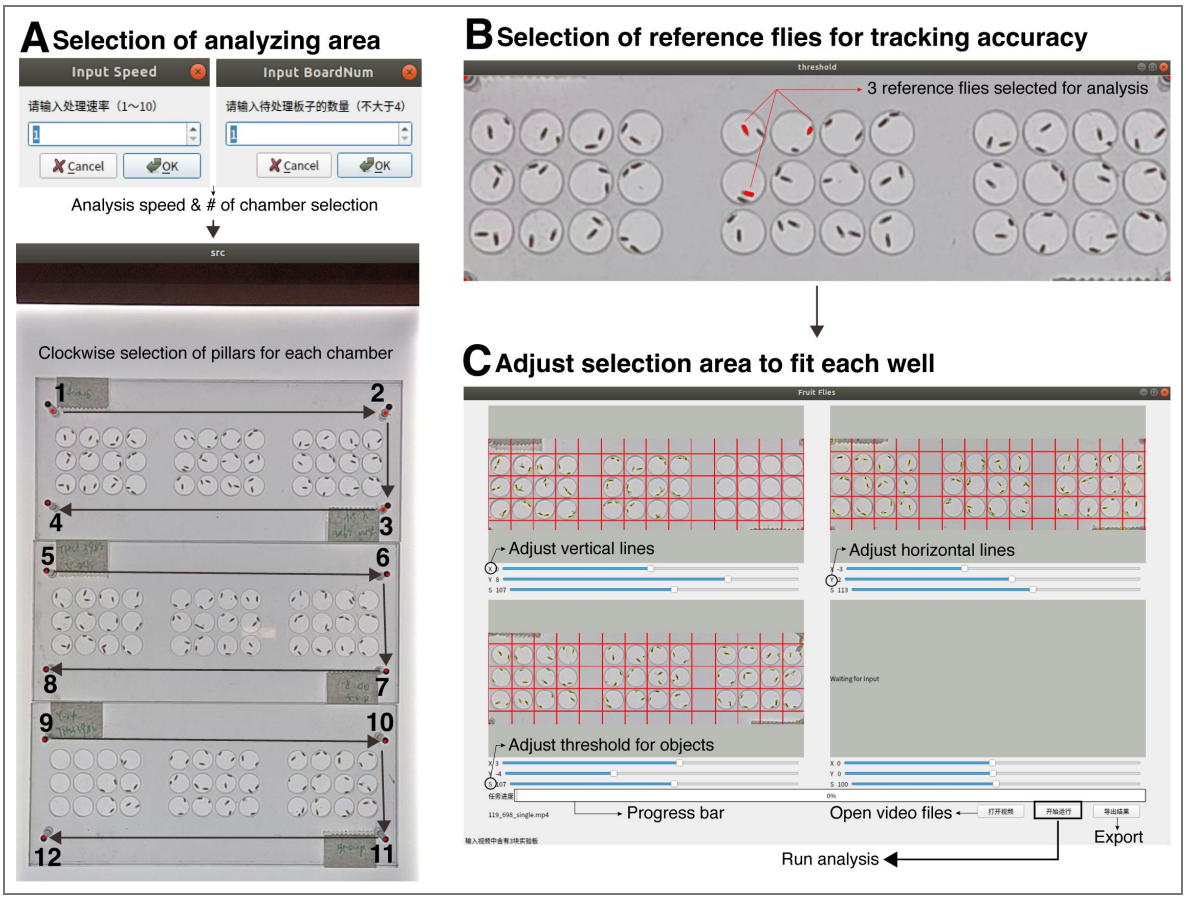

Figure S2. Usage of Linux version of DrosoMating

(A) Typing input speed and input board numbers before selecting analyzing area (upper). Schematic representation of the region selection process (lower). Columns should be selected in a clockwise direction starting from the top left corner. **(B)** Selection of reference flies for tracking accuracy. Any three reference flies were chosen for the analysis. **(C)** The DrosoMating home interface. In brief, “Open videos” to import the recorded mating video. “Run analysis” to initiate the analysis process. The progress bar indicates the status of the analysis. Export raw data by clicking “Export”.

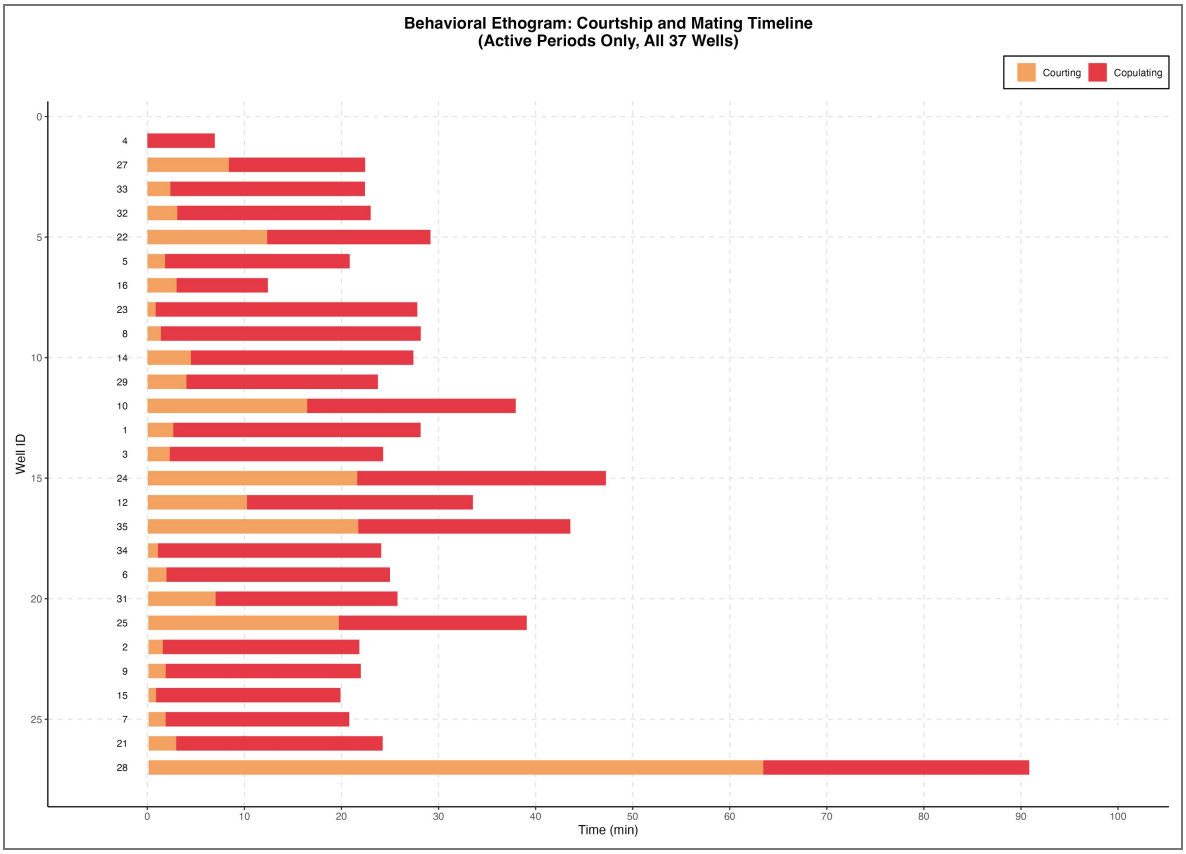

Figure S3. Behavioral ethogram of courtship and copulation dynamics across individual wells.

Horizontal bars depict the temporal progression of male mating behaviors. Orange indicates courtship, red indicates copulation, and the x-axis represents time in minutes. Each row corresponds to a single well labeled with well ID on the left. Only wells with successful copulation are shown, sorted by courtship onset.

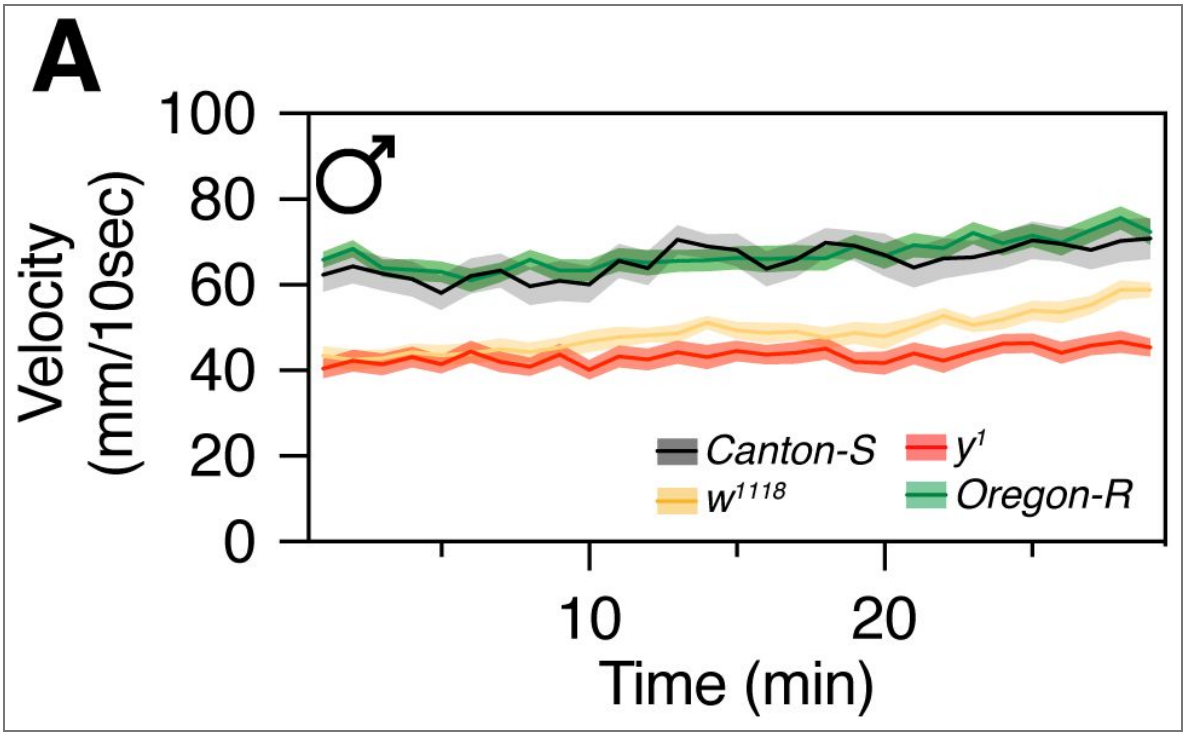

Figure S4. **Basal locomotor activity of male flies from different strains.**

(A) Average locomotor velocity of single-housed male flies from Canton-S, Oregon-R, w^{1118}, and y^{1} strains measured in the absence of females. Velocity was quantified over time and plotted as mm per 10 s. Compared with Canton-S and Oregon-R controls, w^{1118} and y^{1} males exhibited reduced basal locomotor activity, indicating that general motor activity contributes to strain-dependent behavioral differences observed in mating assays.

Data availability

Strains are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables. Our software is freely available at GitHub *https://github.com/hcls-kimlab/DrosoMating*. (Linux version: *https://github.com/liuxiao916/Fruit_Flies*)

Acknowledgements

The authors thank Kim lab members for their kind advice on this manuscript. This work was supported by Startup funds from The HIT Center for Life Science (HCLS) to Woo Jae Kim.

Additional information

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Woo Jae Kim (*wkim@hit.edu.cn*)

Materials availability

This study did not generate new, unique reagents.

Author information

Authors' contributions

XL, YS, and DS conceived and designed the study. WK designed the chamber. XL and FJ completed the development of the software underlying code. YS, HM conducted the subsequent software development and testing. YX fabricated all the hardware devices. YS performed the DrosoMating analysis. WK, YS, HM, and DS wrote the manuscript. WK, YS, HM and DS created the figures and tutorial video. YS and HM revised the manuscript. All authors read, reviewed, and approved the final manuscript.

Funding

Funder	Grant reference number	Author
Harbin Institute of Technology (HIT)	Startup funds	Woo Jae Kim

Author ORCID iDs

Woo Jae Kim: https://orcid.org/0000-0002-2826-4177

References

Barwell T, Raina S, Seroude L (2021) Versatile method to measure locomotion in adult Drosophila1. *Genome* **64**:139-145 https://doi.org/10.1139/gen-2020-0044 | PubMed

Benton R (2015) Neural Circuits: Male Mating Motifs. *Neuron* **87**:912-914 https://doi.org/10.1016/j.neuron.2015.08.017 | PubMed

Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, Schiegg M, Ales J, Beier T, Rudy M, *et al.* (2019) . ilastik: interactive machine learning for (bio)image analysis. *Nat Methods* **16**:1226-1232 https://doi.org/10.1038/s41592-019-0582-9 | PubMed

Branson K, Robie AA, Bender J, Perona P, Dickinson MH (2009) High-throughput ethomics in large groups of Drosophila. *Nat Methods* **6**:451-457 https://doi.org/10.1038/nmeth.1328 | PubMed

Bretman A, Fricke C, Chapman T (2009) Plastic responses of male Drosophila melanogaster to the level of sperm competition increase male reproductive fitness. *Proc Royal Soc B Biological Sci* **276**:1705-1711 https://doi.org/10.1098/rspb.2008.1878 | PubMed

Bretman A, Westmancoat JD, Gage MJG, Chapman T (2011) Males Use Multiple, Redundant Cues to Detect Mating Rivals. *Curr Biol* **21**:617-622 https://doi.org/10.1016/j.cub.2011.03.008 | PubMed

Chen C-H, Lin Y-C, Wang S-H, Kuo T-H, Tsai H-Y (2024a) An automatic system for recognizing fly courtship patterns via an image processing method. *Behav Brain Funct* **20**:5 https://doi.org/10.1186/s12993-024-00231-4 | PubMed

Claridge-Chang A, Assam PN (2016) Estimation statistics should replace significance testing. *Nat Methods* **13**:108-109 https://doi.org/10.1038/nmeth.3729 | PubMed

Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P (2009b) Automated monitoring and analysis of social behavior in Drosophila. *Nat Methods* **6**:297-303 https://doi.org/10.1038/nmeth.1310 | PubMed

Dauwalder B (2008) Systems Behavior: Of Male Courtship, the Nervous System and Beyond in Drosophila. *Curr Genom* **9**:517-524 https://doi.org/10.2174/138920208786847980 | PubMed

Drapeau MD, Cyran SA, Viering MM, Geyer PK, Long AD (2006) A cis-regulatory Sequence Within the yellow Locus of Drosophila melanogaster Required for Normal Male Mating Success. *Genetics* **172**:1009-1030 https://doi.org/10.1534/genetics.105.045666 | PubMed

Dukas R (2020) Natural history of social and sexual behavior in fruit flies. *Sci Rep* **10**:21932 https://doi.org/10.1038/s41598-020-79075-7 | PubMed

Eastwood L, Burnet B (1977a) Courtship latency in maleDrosophila melanogaster. *Behav Genet* **7**:359-372 https://doi.org/10.1007/bf01077449 | PubMed

Ejima A, Griffith LC (2007) Measurement of Courtship Behavior in Drosophila melanogaster. *Cold Spring Harb Protoc* https://doi.org/10.1101/pdb.prot4847

Eyjolfsdottir E, Branson S, Burgos-Artizzu XP, Hoopfer ED, Schor J, Anderson DJ, Perona P (2014) Detecting Social Actions of Fruit Flies. In: Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science. pp. 772-787 https://doi.org/10.1007/978-3-319-10605-2_50

Gal A, Saragosti J, Kronauer DJ (2020) anTraX, a software package for high-throughput video tracking of color-tagged insects. *eLife* **9**:e58145 https://doi.org/10.7554/elife.58145 | PubMed

Gil-Martí B, Barredo CG, Pina-Flores S, Poza-Rodriguez A, Treves G, Rodriguez-Navas C, Camacho L, Pérez-Serna A, Jimenez I, Brazales L, *et al.* (2023a) A simplified courtship conditioning protocol to test learning and memory in Drosophila. *Star Protoc* **4**:101572 https://doi.org/10.1016/j.xpro.2022.101572 | PubMed

Greenspan RJ, Ferveur J-F (2000a) Courtship in Drosophila. *Annu Rev Genet* **34**:205-232 https://doi.org/10.1146/annurev.genet.34.1.205

Griffith LC, Ejima A (2009a) Courtship learning in Drosophila melanogaster: Diverse plasticity of a reproductive behavior. *Learn Mem* **16**:743-750 https://doi.org/10.1101/lm.956309 | PubMed

Hall JC (1994) The Mating of a Fly. *Science* **264**:1702-1714 https://doi.org/10.1126/science.8209251 | PubMed

Huang Y, Kwan A, Kim WJ (2024) Y chromosome genes interplay with interval timing in regulating mating duration of male Drosophila melanogaster. *Gene Rep* **101999** https://doi.org/10.1016/j.genrep.2024.101999

Iyengar A, Imoehl J, Ueda A, Nirschl J, Wu C-F (2012) Automated Quantification of Locomotion, Social Interaction, and Mate Preference in Drosophila Mutants. *J Neurogenet* **26**:306-316 https://doi.org/10.3109/01677063.2012.729626 | PubMed

Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson K (2013) JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat Methods* **10**:64-67 https://doi.org/10.1038/nmeth.2281 | PubMed

Kamyshev NG, Iliadi KG, Bragina JV (1999) Drosophila Conditioned Courtship: Two Ways of Testing Memory. *Learn Mem* **6**:1-20 https://doi.org/10.1101/lm.6.1.1 | PubMed

Kang K, Panzano VC, Chang EC, Ni L, Dainis AM, Jenkins AM, Regna K, Muskavitch MAT, Garrity PA (2012) Modulation of TRPA1 thermal sensitivity enables sensory discrimination in Drosophila. *Nature* **481**:76-80 https://doi.org/10.1038/nature10715 | PubMed

Kim WJ, Jan LY, Jan YN (2013a) A PDF/NPF Neuropeptide Signaling Circuitry of Male Drosophila melanogaster Controls Rival-Induced Prolonged Mating. *Neuron* **80**:1190-1205 https://doi.org/10.1016/j.neuron.2013.09.034 | PubMed

Kim WJ, Jan LY, Jan YN (2012a) Contribution of visual and circadian neural circuits to memory for prolonged mating induced by rivals. *Nat Neurosci* **15**:876-883 https://doi.org/10.1038/nn.3104 | PubMed

Kim WJ, Lee SG, Schweizer J, Auge A-C, Jan LY, Jan YN (2016) Sexually experienced male Drosophila melanogaster uses gustatory-to-neuropeptide integrative circuits to reduce time investment for mating. *bioRxiv* 088724 https://doi.org/10.1101/088724

Kim WJ, Song Y, Zhang T, Zhang X, Ryu TH, Wong KC, Wu Z, Wei Y, Schweizer J, Nguyen K-NH, *et al.* (2024) Peptidergic neurons with extensive branching orchestrate the internal states and energy balance of male Drosophila melanogaster. *bioRxiv* 597277 https://doi.org/10.1101/2024.06.04.597277

Kim Y-K (2009) Sexual Selection and Aggressive Behavior in Drosophila. In: Kim Y-K (Ed). *Handbook of Behavior Genetics* Springer. pp. 317-330 https://doi.org/10.1007/978-0-387-76727-7_22

Kitamoto T (2001) Conditional modification of behavior in Drosophila by targeted expression of a temperature-sensitive shibire allele in defined neurons. *J Neurobiol* **47**:81-92 https://doi.org/10.1002/neu.1018 | PubMed

Krstic D, Boll W, Noll M (2013) Influence of the White Locus on the Courtship Behavior of Drosophila Males. *PLoS ONE* **8**:e77904 https://doi.org/10.1371/journal.pone.0077904 | PubMed

Lee SG, Kang C, Saad B, Nguyen K-NH, Guerra-Phalen A, Bui D, Abbas A-H, Trinh B, Malik A, Zeghal M, *et al.* (2022) Multisensory inputs control the regulation of time investment for mating by sexual experience in male Drosophila melanogaster. *bioRxiv* https://doi.org/10.1101/2022.09.23.509131

Lee SG, Sun D, Miao H, Wu Z, Kang C, Saad B, Nguyen K-NH, Guerra-Phalen A, Bui D, Abbas A-H, *et al.* (2023a) Taste and pheromonal inputs govern the regulation of time investment for mating by sexual experience in male Drosophila melanogaster. *PLOS Genet* **19**:e1010753 https://doi.org/10.1371/journal.pgen.1010753 | PubMed

Levin LR, Han P-L, Hwang PM, Feinstein PG, Davis RL, Reed RR (1992) The Drosophila learning and memory gene rutabaga encodes a Ca2+calmodulin-responsive adenylyl cyclase. *Cell* **68**:479-489 https://doi.org/10.1016/0092-8674(92)90185-f | PubMed

Liu X (2024) Fruit_Flies. GitHub. https://github.com/hcls-kimlab/DrosoMating

Neckameyer WS, Bhatt P (2016) Drosophila, Methods and Protocols. *Methods Mol Biol* **1478**:303-320 https://doi.org/10.1007/978-1-4939-6371-3_19 | PubMed

Pavlou HJ, Goodwin SF (2013) Courtship behavior in Drosophila melanogaster: towards a 'courtship connectome.'. *Curr Opin Neurobiol* **23**:76-83 https://doi.org/10.1016/j.conb.2012.09.002 | PubMed

Pérez-Escudero A, Vicente-Page J, Hinz RC, Arganda S, Polavieja GG de (2014) idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nat Methods* **11**:743-748 https://doi.org/10.1038/nmeth.2994 | PubMed

Qu S, Zhu Q, Zhou H, Gao Y, Wei Y, Ma Y, Wang Z, Sun X, Lei Zhang, Yang Q, *et al.* (2022) EasyFlyTracker: A Simple Video Tracking Python Package for Analyzing Adult Drosophila Locomotor and Sleep Activity to Facilitate Revealing the Effect of Psychiatric Drugs. *Front Behav Neurosci* **15**:809665 https://doi.org/10.3389/fnbeh.2021.809665 | PubMed

Reza MdA, Mhatre SD, Morrison JC, Utreja S, Saunders AJ, Breen DE, Marenda DR (2013) Automated analysis of courtship suppression learning and memory in Drosophila melanogaster. *Fly* **7**:105-111 https://doi.org/10.4161/fly.24110 | PubMed

Risse B, Berh D, Otto N, Klämbt C, Jiang X (2017) FIMTrack: An open source tracking and locomotion analysis software for small animals. *PLoS Comput Biol* **13**:e1005530 https://doi.org/10.1371/journal.pcbi.1005530 | PubMed

Risse B, Otto N, Berh D, Jiang X, Klämbt C (2014) FIM Imaging and FIMtrack: Two New Tools Allowing High-throughput and Cost Effective Locomotion Analysis. *J Vis Exp* https://doi.org/10.3791/52207-v

Rodriguez A, Zhang H, Klaminder J, Brodin T, Andersson PL, Andersson M (2017) ToxTrac_: A fast and robust software for tracking organisms. *Methods Ecol Evol* **9**:460-464 https://doi.org/10.1111/2041-210x.12874

Simon AF, Chou M-T, Salazar ED, Nicholson T, Saini N, Metchev S, Krantz DE (2011) A simple assay to study social behavior in Drosophila: measurement of social space within a group. Genes, brain. *Behav* **11**:243-52 https://doi.org/10.1111/j.1601-183x.2011.00740.x | PubMed

Singh A, Singh BN (2014) Mating latency, duration of copulation and fertility in four species of the Drosophila bipectinata complex. *Indian J Exp Biol* **52**:175-80 PubMed

Singh BN, Singh A (2016) The genetics of sexual behavior in Drosophila. *Adv Genom Genet* **6**:1-9 https://doi.org/10.2147/agg.s58525

Sokolowski MB (2001) Drosophila: Genetics meets behaviour. *Nat Rev Genet* **2**:879-890 https://doi.org/10.1038/35098592 | PubMed

Stern U, Zhu EY, He R, Yang C-H (2015) Long-duration animal tracking in difficult lighting conditions. *Sci Rep* **5**:10432 https://doi.org/10.1038/srep10432 | PubMed

Sun D, Zhang X, Miao H, Zhang T, Kim WJ (2024) Molecular and Neural Circuit Mechanisms Underlying Sexual Experience-dependent Long-Term Memory in Drosophila. *bioRxiv* https://doi.org/10.1101/2024.09.28.615582

Sun Y, Zhang X, Wu Z, Li W, Kim WJ (2024) Genetic Screening Reveals Cone Cell-Specific Factors as Common Genetic Targets Modulating Rival-Induced Prolonged Mating in male Drosophila melanogaster. *G3: Genes, Genomes, Genet* jkae255 https://doi.org/10.1093/g3journal/jkae255

Vosshall LB (2007) Into the mind of a fly. *Nature* **450**:193-197 https://doi.org/10.1038/nature06335 | PubMed

Wolf R, Wittig T, Liu L, Wustmann G, Eyding D, Heisenberg M (1998) Drosophila Mushroom Bodies Are Dispensable for Visual, Tactile, and Motor Learning. *Learn Mem* **5**:166-178 https://doi.org/10.1101/lm.5.1.166 | PubMed

Wong K, Schweizer J, Nguyen K-NH, Atieh S, Kim WJ (2019) Neuropeptide relay between SIFa signaling controls the experience-dependent mating duration of male Drosophila. *bioRxiv* https://doi.org/10.1101/819045

Yamamoto D, Koganezawa M (2013) Genes and circuits of courtship behaviour in Drosophila males. *Nat Rev Neurosci* **14**:681-692 https://doi.org/10.1038/nrn3567 | PubMed

Yamanaka O, Takeuchi R (2018) UMATracker: an intuitive image-based tracking platform. *J Exp Biol* **221**:jeb182469 https://doi.org/10.1242/jeb.182469 | PubMed

Yapici N, Kim Y-J, Ribeiro C, Dickson BJ (2008) A receptor that mediates the post-mating switch in Drosophila reproductive behaviour. *Nature* **451**:33-37 https://doi.org/10.1038/nature06483 | PubMed

Zhang T, Wu Z, Song Y, Li W, Sun Y, Zhang X, Wong K, Schweizer J, Nguyen K-NH, Kwan A, *et al.* (2024a) Long-range neuropeptide relay as a central-peripheral communication mechanism for the context-dependent modulation of interval timing behaviors. *bioRxiv* https://doi.org/10.1101/2024.06.03.597273

Zhang T, Zhang X, Sun D, Kim WJ (2024b) Exploring the Asymmetric Body's Influence on Interval Timing Behaviors of Drosophila melanogaster. *Behav Genet* 1-10 https://doi.org/10.1007/s10519-024-10193-y | PubMed

Zhang X, Miao H, Kang D, Sun D, Kim WJ (2024) Male-specific sNPF peptidergic circuits control energy balance for mating duration through neuron-glia interactions. *bioRxiv* https://doi.org/10.1101/2024.10.17.618859

Peer reviews

Reviewer #1 (Public review):

Summary:

The study of Drosophila mating behaviors has offered a powerful entry point for understanding how complex innate behaviors are instantiated in the brain. The effectiveness of this behavioral model stems from how readily quantifiable many components of the courtship ritual are, facilitating the fine-scale correlations between the behaviors and the circuits that underpin their implementation. Detailed quantification, however, can be both time consuming and error prone, particularly when scored manually. Song et al. have sought to address this challenge by developing DrosoMating, software that facilitates the automated and high-throughput quantification of 6 common metrics of courtship and mating behaviors. Compared to a human observer, DrosoMating matches courtship scoring with high fidelity. Further, the authors demonstrate that the software effectively detects previously described variations in courtship resulting from genetic background or social conditioning. Finally, they validate its utility in assaying the consequences of neural manipulations by silencing Kenyon cells involved in memory formation in the context of courtship conditioning.

Strengths:

(1) The authors demonstrate that for three key courtship/mating metrics, DrosoMating performs virtually indistinguishably from a human observer, with differences consistently within 10 seconds and no statistically significant differences detected. This demonstrates the software's usefulness as a tool for reducing bias and scoring time for analyses involving these metrics.

(2) The authors validate the tool across multiple genetic backgrounds and experimental manipulations to confirm its ability to detect known influences on male mating behavior.

(3) The authors present a simple, modular chamber design that is integrated with DrosoMating and allows for high throughput experimentation, capable of simultaneously analyzing up to 144 fly pairs across all chambers.

Weaknesses:

(1) DrosoMating appears to be an effective tool for the quantification of key courtship and mating metrics, but similar tools for automated analysis already exist. The authors present a compelling use case for DrosoMating, where it has particular advantages over tools like FlyTracker and Ctrax for high-throughput analysis. This comparative analysis, however, leaves out modern pose-estimation methods (SLEAP, DeepLabCut), better able to tolerate low contrast and occlusion. It therefore remains unclear what specific advantages it might offer over current machine learning approaches.

(2) The courtship behaviors of Drosophila males represent a series of complex behaviors that unfold dynamically in response to female signals. While metrics like courtship latency, courtship index, and mating duration are useful, they compress the complexity of actions that occur throughout the mating ritual. The authors suggest DrosoMating's modular architecture facilitates integration with behavioral classifiers like JAABA. Such integration could substantially expand the utility of this tool for the broader Drosophila neuroscience community, but in its current form its applications are confined to summary timing metrics.

(3) Validation is limited to multiple D. melanogaster strains. Cross-species studies of mating behavior diversity are increasingly common, so demonstrating the tool's accuracy across species would strengthen claims about its broader applicability.

https://doi.org/10.7554/eLife.109742.2.sa2

Reviewer #2 (Public review):

This manuscript introduces DrosoMating, an integrated hardware-software pipeline designed to automate quantification of Drosophila courtship and mating behavior. The authors aim to provide a low-cost, scalable alternative to existing behavioral tracking systems, focusing on extracting key temporal metrics including courtship index, copulation latency, and mating duration from high-throughput video recordings.

A major strength of the work is the clear emphasis on experimental scalability and practical usability. The system is designed for multi-chamber recording and performs robustly under low-quality imaging conditions where conventional pose-tracking pipelines often fail. The revised manuscript substantially improves its scientific positioning through the inclusion of systematic benchmarking against widely used tools (Ctrax and FlyTracker), demonstrating that both fail to complete end-to-end analysis under these recording conditions: Ctrax due to segmentation instability and trajectory fragmentation, FlyTracker due to runtime errors during feature computation. A comparative table (Table 1) summarizes key features across tools, and the authors appropriately qualify that these limitations are specific to the low-quality video conditions tested and should not be interpreted as general shortcomings of those tools. The addition of individual-level behavioral ethograms (Figure S3) further strengthens the evidence by allowing direct assessment of the system's temporal resolution at the single-fly level.

The authors also appropriately address statistical concerns raised in review. The re-analysis using ANOVA frameworks - one-way ANOVA with Tukey's HSD for multi-strain comparisons, two-way ANOVA with Sidak's correction for genotype × training interactions - improves the rigor of the behavioral comparisons and supports the revised conclusions regarding strain and learning effects. The addition of locomotor control analyses (Figure S4) further clarifies interpretation of mutant phenotypes by partially disentangling motor from courtship-specific effects, with the revised text appropriately acknowledging contributions from both general hypoactivity and sensory impairments.

A key limitation remains the conceptual scope of the system. DrosoMating is optimized for state-based temporal segmentation rather than fine-grained behavioral annotation or posture-level decomposition. While the authors now clearly acknowledge this and position the tool appropriately, it inherently restricts its use cases compared to modern pose-estimation and classifier-based frameworks. Additionally, the benchmarking comparison is necessarily asymmetric: DrosoMating is tested on low-quality videos where it excels by design, while Ctrax and FlyTracker are evaluated under conditions outside their intended operating range. A comparison under more favorable conditions for the tracking-based tools, or an evaluation of whether modest improvements in video quality would bring conventional pipelines within functional range, would further contextualize the practical boundary between approaches. The comparison also does not include modern deep-learning-based tools (e.g., DeepLabCut, SLEAP), which may be more robust to low-contrast conditions than classical segmentation-based pipelines. These points do not diminish the demonstrated utility of DrosoMating for its intended niche but would help users make more informed decisions about tool selection.

Overall, the revised manuscript presents a well-validated and clearly positioned contribution. It defines the niche in which DrosoMating provides substantial practical value while

appropriately delimiting its limitations relative to more general behavioral analysis frameworks.

https://doi.org/10.7554/eLife.109742.2.sa1

Author response:

The following is the authors' response to the original reviews.

Public Reviews:

Reviewer #1 (Public review):

Summary:

The study of Drosophila mating behaviors has offered a powerful entry point for understanding how complex innate behaviors are instantiated in the brain. The effectiveness of this behavioral model stems from how readily quantifiable many components of the courtship ritual are, facilitating the fine-scale correlations between the behaviors and the circuits that underpin their implementation. Detailed quantification, however, can be both time-consuming and error-prone, particularly when scored manually. Song et al. have sought to address this challenge by developing DrosoMating, software that facilitates the automated and high-throughput quantification of 6 common metrics of courtship and mating behaviors. Compared to a human observer, DrosoMating matches courtship scoring with high fidelity. Further, the authors demonstrate that the software effectively detects previously described variations in courtship resulting from genetic background or social conditioning. Finally, they validate its utility in assaying the consequences of neural manipulations by silencing Kenyon cells involved in memory formation in the context of courtship conditioning.

Strengths:

(1) The authors demonstrate that for three key courtship/mating metrics, DrosoMating performs virtually indistinguishably from a human observer, with differences consistently within 10 seconds and no statistically significant differences detected. This demonstrates the software's usefulness as a tool for reducing bias and scoring time for analyses involving these metrics.

(2) The authors validate the tool across multiple genetic backgrounds and experimental manipulations to confirm its ability to detect known influences on male mating behavior.

(3) The authors present a simple, modular chamber design that is integrated with DrosoMating and allows for high-throughput experimentation, capable of simultaneously analyzing up to 144 fly pairs across all chambers.

Weaknesses:

(1) DrosoMating appears to be an effective tool for the high-throughput quantification of key courtship and mating metrics, but a number of similar tools for automated analysis already exist. FlyTracker (CalTech), for instance, is a widely used software that offers a similar machine vision approach to quantifying a variety of courtship metrics. It would be valuable to understand how DrosoMating compares to such approaches and what specific advantages it might offer in terms of accuracy, ease of use, and sensitivity to experimental conditions.

(2) The courtship behaviors of Drosophila males represent a series of complex behaviors that unfold dynamically in response to female signals (Coen et al., 2014; Ning et al., 2022; Roemschied et al., 2023). While metrics like courtship latency, courtship index, and

copulation duration are useful summary statistics, they compress the complexity of actions that occur throughout the mating ritual. The manuscript would be strengthened by a discussion of the potential for DrosoMating to capture more of the moment-to-moment behaviors that constitute courtship. Even without modifying the software, it would be useful to see how the data can be used in combination with machine learning classifiers like JAABA to better segment the behavioral composition of courtship and mating across genotypes and experimental manipulations. Such integration could substantially expand the utility of this tool for the broader Drosophila neuroscience community.

(3) While testing the software's capacity to function across strains is useful, it does not address the "universality" of this method. Cross-species studies of mating behavior diversity are becoming increasingly common, and it would be beneficial to know if this tool can maintain its accuracy in Drosophila species with a greater range of morphological and behavioral variation. Demonstrating the software's performance across species would strengthen claims about its broader applicability.

Reviewer #2 (Public review):

This paper introduces "DrosoMating," an integrated hardware and software solution for automating the analysis of male Drosophila courtship. The authors aim to provide a low-cost, accessible alternative to expensive ethological rigs by utilizing a custom acrylic chamber and smartphone-based recording. The system focuses on quantifying key temporal metrics-Courtship Index (CI), Copulation Latency (CL), and Mating Duration (MD)-and is applied to behavioral paradigms involving memory mutants (orb2, rut).

The development of open-source behavioral tools is a significant contribution to neuroethology, and the authors successfully demonstrate a system that simplifies the setup for large-scale screens. A major strength of the work is the specific focus on automating Copulation Latency and Mating Duration, metrics that are often labor-intensive to score manually.

However, there are several limitations in the current analysis and validation that affect the strength of the conclusions:

First, the statistical rigor requires substantial improvement. The analysis of multi-group experiments (e.g., comparing four distinct strains or factorial designs with genotype and training) currently relies on multiple independent Student's t-tests. This approach is statistically invalid for these experimental designs as it inflates the family-wise Type I error rate. To support the claims of strain-specific differences or learning deficits, the data must be analyzed using Analysis of Variance (ANOVA) to properly account for multiple comparisons and to explicitly test for interaction effects between genotype and training conditions.

Second, the biological validation using w^{1118} and y^1 mutants entails a potential confound. The authors attribute the low Courtship Index in these strains to courtship-specific deficits. However, both strains are known to exhibit general locomotor sluggishness (due to visual or pigmentation/behavioral defects). Since "following" behavior is likely a component of the Courtship Index, a reduction in this metric could reflect a general motor deficit rather than a specific lack of reproductive motivation. Without controlling for general locomotion, the interpretation of these behavioral phenotypes remains ambiguous.

Third, the benchmarking of the system is currently limited to comparisons against manual scoring. Given that the field has largely adopted sophisticated open-source tracking tools (e.g., Ctrax, FlyTracker, JAABA), the utility of DrosoMating would be better

> *contextualized by comparing its performance - in terms of accuracy, speed, or identity maintenance - against these existing automated standards, rather than solely against human observation.*
>
> *Finally, the visual presentation of the data hinders the assessment of the system's temporal precision. While the system is designed to capture time-resolved metrics, the results are presented primarily as aggregate bar plots. The absence of behavioral ethograms or raster plots makes it difficult to verify the software's ability to accurately detect specific transitions, such as the exact onset of copulation.*

We sincerely thank the reviewers for their constructive and detailed feedback, which has substantially improved the clarity and rigor of our work. Below is a summary of the major revisions.

(1) Comparison with existing tools. We added a new main figure (Figure 5) and Table 1 systematically benchmarking DrosoMating against Ctrax and FlyTracker on identical low-quality, high-throughput mating videos. Both conventional tools failed under our recording conditions: Ctrax showed severe segmentation instability and fragmented trajectories, while FlyTracker frequently crashed during feature computation. These results demonstrate that DrosoMating's state-detection approach bypasses the pose-tracking limitations that impair established pipelines. We also revised the Discussion to clarify that DrosoMating is specialized for robust extraction of mating timing metrics from low-quality videos, not a replacement for general-purpose tracking tools.

(2) Statistical analysis. Following Reviewer #2's recommendation, we re-analyzed Figure 3 using one-way ANOVA with Tukey's HSD for strain comparisons, and Figure 4 using two-way ANOVA with Sidak's post-hoc test for learning assays, including the critical Genotype by Training interaction term.

(3) New supplementary data. We added Figure S4 showing basal locomotor velocity of single-housed males across all four strains to decouple motor defects from courtship deficits in w1118 and y1 mutants. We also added Figure S3 with behavioral ethograms for individual flies to visually demonstrate the system's temporal resolution and detection accuracy.

(4) Additional improvements. We standardized statistical reporting across all figure legends, explicitly defined the segmentation threshold parameter s in the Methods, consistently used "Mating Duration (MD)" throughout, standardized MB247-GAL4 labeling, clarified that occluded frames are retained for mating-state detection via merged-contour analysis, and corrected minor errors including duplicate references and unnecessary quotation marks.

We believe these revisions substantially strengthen the manuscript and clearly position DrosoMating within the existing ecosystem of behavioral analysis tools.

We remain grateful for the valuable feedback from both reviewers and the editorial team, and we hope the revised version meets the standards for publication in eLife.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations for the authors):***
>
> *(1) It's difficult to assess the utility of this tool in relation to the variety of alternative methods available for the automated scoring of courtship behavior, some of which offer more granular behavioral data than what DrosoMating has been presented to produce. A direct comparison to other approaches (e.g., FlyTracker, DeepLabCut, SLEAP) would be highly valuable for understanding what use cases DrosoMating is best suited for. Ideally, this analysis would include (i) a comparison of the accuracy of scoring courtship metrics and constituent behaviors, (ii) an assessment of the sensitivity of the various software to*

> *different experimental conditions, such as lighting and alternative chamber designs, and (iii) testing across Drosophila species, which could help highlight the particular strengths of DrosoMating over more established methods. At a minimum, a table comparing key features, requirements, and capabilities would help position DrosoMating within the existing ecosystem of tools.*

We sincerely appreciate the reviewer's insightful comments on benchmarking DrosoMating against existing tools for automated courtship behavior analysis. We fully agree that direct comparisons are essential to clarify the unique advantages and ideal use cases of DrosoMating. We have extensively revised the manuscript by adding comparative experiments, a new figure（Figure 5) and comparative table (Table1), and expanded descriptions, as detailed below.

(1) Direct comparison with conventional tracking tools

We systematically tested Ctrax and FlyTracker on the same low-quality, high-throughput mating videos used for DrosoMating. The corresponding results are presented in Figure 5.

Ctrax showed severe segmentation instability, including over-segmentation, under-segmentation, and fragmented trajectories, and failed to stably detect two flies.

FlyTracker was able to generate background models and showed partially improved segmentation after threshold adjustment, but frequently crashed during feature computation and could not complete the end-to-end analysis pipeline.

These results confirm that tracking-based tools are vulnerable to low contrast, chamber artifacts, and prolonged male–female overlap, whereas DrosoMating bypasses pose tracking and directly detects mating states.

The revised manuscript text is as follows (line 265-342):

"DrosoMating is more compatible with low-quality mating videos than conventional tracking-based pipelines.

To evaluate whether conventional fly-tracking pipelines could be used as alternative tools for extracting mating-duration metrics, we tested Ctrax and FlyTracker on representative low-quality single-chamber mating videos recorded under our standard high-throughput conditions (Fig. 5). Ctrax is designed to estimate the position and orientation of multiple walking flies while maintaining individual identities over time (Branson et al., 2009) (*https://ctrax.sourceforge.net/*), whereas FlyTracker aims to track fly pose, including position, orientation, body size, wing and leg positions, and to generate trajectory- and feature-based outputs for downstream behavior analysis (Eyjolfsdottir et al., 2014). Because these programs were not readily compatible with our full high-throughput behavioral recording setup, we first cropped the original videos and tested single-chamber videos containing one male and one female fly.

Using Ctrax, we observed that target detection was highly variable even within single-chamber videos. In representative frames, Ctrax could occasionally identify the two flies correctly (Fig. 5A). However, imperfect segmentation was frequently observed. In some frames, parts of the fly body were detected as additional targets, resulting in over-segmentation and an apparent increase in the number of detected flies (Fig. 5B, C). Conversely, when the male and female were close to each other or physically overlapped, the two animals were sometimes detected as a single target (Fig. 5D). These examples indicate that Ctrax detection was sensitive to the low contrast and overlapping fly bodies present in our mating videos. We then adjusted the Ctrax detection threshold to improve segmentation quality. Although threshold optimization improved detection in some frames, abnormal detections remained evident across randomly sampled frames (Fig. 5E). For example, some

frames still showed incorrect target numbers, and a severe segmentation failure was observed in the lower-right example, where the detected objects did not correspond to two clearly separable flies. Thus, even after parameter optimization, Ctrax did not consistently maintain the expected two-target detection state in single-chamber mating videos.

The instability of Ctrax detection was also reflected in the trajectory output. In the first 500 frames, the generated trajectories were fragmented into multiple colored track segments rather than two continuous trajectories corresponding to the male and female (Fig. 5F). In addition, some trajectories extended outside the chamber boundary, indicating tracking errors and identity instability. Consistently, frame-by-frame quantification of detected target number showed that the detected object count did not remain stable at the expected value of two flies per chamber (Fig. 5G). Because Ctrax failed to maintain stable two-fly detection and continuous trajectories under these video conditions, its output could not be reliably used for downstream extraction of copulation latency or mating duration.

We next tested FlyTracker on the same type of cropped single-chamber mating videos. FlyTracker was developed to track multiple flies by estimating body position, orientation, size, wing and leg positions, and by maintaining fly identities across video frames; it also outputs per-frame features such as velocity, facing angle, and wing-angle-related measurements for downstream behavioral analysis (Eyjolfsdottir et al., 2014). In our videos, FlyTracker was able to generate a background model, indicating that the program could recognize the overall imaging field and chamber background (Fig. 5H). However, during calibration and segmentation, the default threshold setting produced inconsistent detection results (Fig. 5I). Although some frames were segmented relatively well under the default threshold (Fig. 5J), these successful examples were not representative of the overall tracking process, and the program frequently terminated with runtime errors during tracking or downstream feature computation. To improve detection stability, we lowered the segmentation threshold. Under this adjusted setting, FlyTracker produced more complete fly masks in representative frames (Fig. 5K), and the diagnostic output showed that two flies were detected in many sampled frames (Fig. 5L). Nevertheless, detection remained unstable in some sampled frames, including frames in which zero or one fly was detected despite the expected two flies per chamber (Fig. 5L). The full pipeline ultimately failed during feature computation, producing a runtime error before complete tracking and feature outputs could be generated (Fig. 5M). Therefore, even after threshold adjustment, FlyTracker could not provide a stable end-to-end workflow for extracting mating-duration metrics from these low-quality mating videos.

This failure mode is relevant because FlyTracker depends on stable segmentation, identity maintenance, and per-frame feature extraction. In our assay videos, low contrast, chamber-edge artifacts, and prolonged male–female overlap during copulation interfered with these requirements. As a result, FlyTracker could occasionally identify the flies in individual frames, but it did not reliably complete the full analysis pipeline required for downstream behavioral quantification. This limitation is especially important for workflows such as JAABA, which use manually labeled examples to train behavior classifiers but still depend on upstream tracking-derived features. Thus, for our low-quality high-throughput mating recordings, FlyTracker-based analysis was substantially less practical than DrosoMating, which directly outputs mating-related timing metrics without requiring continuous high-fidelity two-fly pose tracking."

(2) Evaluation of accuracy, robustness, and cross-species potential

We addressed the three key points requested by the reviewer:

(i) Accuracy: DrosoMating achieved 98–99% agreement with manual scoring. Under our experimental conditions, neither Ctrax nor FlyTracker completed an end-to-end workflow capable of reliably extracting copulation latency and mating duration. Ctrax produced

fragmented trajectories with unstable target counts, and FlyTracker terminated with runtime errors during feature computation.

(ii) Robustness: DrosoMating is highly robust to low-contrast lighting and common behavioral chamber setups, whereas conventional tools require high-quality videos and fail under fly occlusion.

(iii) Cross-species testing: We did not perform cross-species validation in this revision. DrosoMating relies on mating state detection rather than species-specific morphology, suggesting potential transferability, although this remains to be experimentally validated. We have therefore revised the text to avoid claiming demonstrated cross-species performance and now describe this as a potential future application.

The revised manuscript text is as follows (line 572-576):

"Cross-species testing was not performed in this study. Although DrosoMating's state-detection approach is morphology-agnostic and therefore theoretically applicable across Drosophila species, we have revised the text to avoid claiming demonstrated cross-species performance. Formal validation across diverse species remains a promising future direction."

We added Table1 to compare key features across DrosoMating, Ctrax, and FlyTracker, including low-quality video compatibility, multi-chamber support, robustness to fly overlap, direct output of mating timing metrics, accuracy.

(3) Clarification of ideal use cases

We revised the Discussion to emphasize that DrosoMating is not intended to replace general-purpose pose or tracking tools (e.g., FlyTracker, Ctrax) that provide fine-grained behavioral data.

Instead, it offers a specialized, robust, and high-throughput workflow for scenarios requiring efficient extraction of mating timing metrics from low-quality videos, where conventional pipelines often fail.

These revisions greatly improve the clarity and positioning of DrosoMating. We thank the reviewer for this valuable suggestion.

The revised manuscript text is as follows (line 578-620):

“Comparative Evaluation with Existing Courtship Analysis Tools

A major advantage of DrosoMating is its compatibility with low-quality, high-throughput mating videos. Conventional tracking-based tools such as Ctrax and FlyTracker are powerful for trajectory- and pose-based behavioral analysis, but they generally require stable object segmentation, identity maintenance, and reliable feature extraction across frames. Ctrax was designed to estimate the positions and orientations of multiple walking flies while maintaining their identities, whereas FlyTracker aims to track detailed fly pose and generate per-frame behavioral features. Under our recording conditions, these requirements were difficult to satisfy because the videos were low contrast and the male and female frequently overlapped during copulation.

In our tests, both tools showed limited compatibility with these videos. Even after cropping to single-chamber videos and adjusting detection parameters, Ctrax produced unstable target numbers, fragmented trajectories, and tracking errors. FlyTracker could generate a background model and occasionally segment flies successfully, but detection remained unstable and the full pipeline failed during feature computation. These issues are particularly relevant for copulation latency and mating duration analysis: during copulation, the male and female remain physically coupled for a long period, which makes identity-based tracking

difficult. For these timing metrics, it is more important to robustly detect the onset and offset of the mating state than to reconstruct detailed individual trajectories. DrosoMating was purpose-built to address these specific challenges. It operates reliably on lower-quality video streams, requires no complex pre-processing or manual ROI definition, and is optimized for high-throughput multi-chamber analysis. While it does not offer the same level of pose or kinematic detail as other tools, it provides a unique solution for laboratories seeking a simple, fast, and robust pipeline to quantify core reproductive timing metrics—copulation latency (CL), courtship index (CI), and mating duration (MD)—without the overhead of more complex systems (Table 1).

Although we directly tested only Ctrax and FlyTracker, this limitation may also affect workflows that depend on upstream tracking-derived features. For example, JAABA uses tracking-derived features to train behavior classifiers, and DANCE, a recent Drosophila aggression and courtship pipeline, uses JAABA-based classifiers and lists FlyTracker and JAABA as required software. Thus, our conclusion is not that these tools are generally unsuitable for Drosophila behavior analysis, but that DrosoMating provides a more practical workflow for low-quality, high-throughput videos focused specifically on mating timing.

While DrosoMating currently prioritizes mating timing metrics over discrete behavioral classification, its modular architecture provides a foundation for future integration with behavior classifiers such as JAABA. Laboratories requiring granular behavioral elements—such as wing extension or circling—would benefit from an extended pipeline that exports per-frame kinematic features for downstream classifier training. Validating this integration represents a promising future direction to broaden the tool's utility beyond core reproductive timing assays."

We have also added this clarification to the Table 1 legend to avoid overgeneralizing the limitations of Ctrax and FlyTracker (line 786-790):

"DrosoMating was designed to extract mating-related timing metrics from high-throughput videos without requiring continuous two-fly identity tracking. Ctrax and FlyTracker were tested on cropped single-chamber videos from the same recording setup. Their limitations described here refer specifically to these low-quality mating videos and should not be interpreted as general limitations of the tools."

> *(2) The coarse nature of the metrics captured by DrosoMating may limit the usefulness of this tool for many researchers. Consider integrating DrosoMating with one or more behavioral classifiers (e.g., JAABA) and validating its performance to increase its utility across a wider range of uses. Demonstrating the feasibility of this would substantially increase the tool's appeal to those who need access to more discrete behavioral elements.*

We thank the reviewer for this constructive suggestion. We agree that DrosoMating is currently optimized for rapid, high-throughput quantification of core mating timing metrics (CL, CI, and MD) rather than discrete behavioral classification (e.g., wing extension, circling, or aggressive postures). This reflects a deliberate design trade-off: by prioritizing robust state detection over detailed pose tracking, DrosoMating achieves reliable performance on low-quality videos where conventional identity-based pipelines fail.

We appreciate the reviewer's vision for expanding the tool's utility. While DrosoMating does not presently generate the per-frame kinematic features (position, orientation, wing angles, etc.) required as input for JAABA classifiers, its modular architecture and underlying video-processing framework provide a foundation for future integration. In the revised Discussion, we have clarified that extending DrosoMating to export trajectory-derived features compatible with downstream classifiers such as JAABA represents a promising future

direction—one that would broaden its applicability to laboratories requiring granular behavioral elements, without necessitating a complete overhaul of the existing pipeline.

We hope this clarification addresses the reviewer's concern and accurately reflects both the current capabilities and future potential of the tool.

The revised manuscript text is as follows (line 614-620):

"While DrosoMating currently prioritizes mating timing metrics over discrete behavioral classification, its modular architecture provides a foundation for future integration with behavior classifiers such as JAABA. Laboratories requiring granular behavioral elements—such as wing extension or circling—would benefit from an extended pipeline that exports per-frame kinematic features for downstream classifier training. Validating this integration represents a promising future direction to broaden the tool's utility beyond core reproductive timing assays."

> *(3) Please clarify how DrosoMating handles fly identification during mating? I would think mounted flies would be occluded, and if these frames are excluded from analysis, it would be expected to skew mating duration scores. It would be helpful if the authors could discuss these details.*

Occluded frames are excluded from CI calculation because individual courtship actions cannot be reliably assigned during prolonged overlap. However, these frames are not discarded from mating-duration analysis. Instead, prolonged merged contours are used as evidence for copulation-state detection, and the MD timer continues until physical separation:

(1) When two flies are separate, the system detects two contours; when they overlap during mounting, it detects one merged contour. Our code processes both cases, so occluded frames are retained, not skipped.

(2) To distinguish a single fly from two overlapping flies, we analyze the shape of the merged contour. Two overlapping flies produce a characteristically different aspect ratio (elongated shape) compared to one fly. This geometric cue is fed into the state classifier to label the frame as "mating."

(3) Because these frames are classified as mating rather than excluded, the mating duration timer runs continuously through the occlusion period. Thus, mating duration is not artificially shortened.

The high agreement between DrosoMating and manual scoring for mating duration (within 10 s, no significant difference) confirms that this approach does not introduce bias.

We revised the manuscript as follow (line 191-193):

"Occluded frames are excluded from CI calculation but retained for mating-state detection via merged-contour analysis, ensuring continuous MD measurement through copulation."

> *(4) Please indicate the statistical tests used in Figures 3, 4, and S1. What methods were used to address multiple hypothesis testing?*

We thank the reviewer for this important comment. For comparisons between two groups, two-sided Student's t-tests were used. For comparisons among three or more groups, one-way ANOVA followed by post-hoc tests were applied. For two-factor experimental designs, two-way ANOVA was used. We have clearly stated these statistical tests in the Statistical Analysis section and have added this information to the figure legends for Figures 3, 4, and S1 in the revised manuscript.

The revised "Statistical Analysis" section is as follows (line 512-524):

"To ensure robust statistical analysis, each experimental group included at least 100 male flies (naïve, sexually experienced, or singly reared). Internal controls were incorporated into every experiment as recommended by Bretman et al. (2011) (Bretman et al., 2011). Normality of the mating duration data was confirmed using the Kolmogorov-Smirnov test ($p>0.05$). For group comparisons, two-sided Student's t-tests were applied to calculate significance levels (****$p<0.0001$, ***$p<0.001$, **$p<0.01$, * $p<0.05$), Comparisons among three or more groups were performed using one-way ANOVA with Tukey's HSD post-hoc tests. Two-factor experimental comparisons were performed using two-way ANOVA followed by Sidak's post-hoc test. while estimation statistics (Claridge-Chang and Assam, 2016) were used to visualize effect sizes, mean differences, and precision, avoiding reliance solely on null hypothesis testing. All analyses, including data plotting, were performed using GraphPad Prism software."

> *(5) Line 43: "High resolution video tracking enables...".*

We thank the reviewer for noting this imprecise expression. We have revised the statement to accurately describe that our method uses image-based video analysis to identify courtship and copulation events and quantify their temporal parameters. The description has been corrected in the revised manuscript line 44-46:

"Our image-based video analysis enables precise identification of courtship and copulation events, as well as quantification of their timing and duration under controlled conditions. "

> *(6) Line 51: remove quotation mark.*

We thank the reviewer for the careful correction. The unnecessary quotation mark at Line 51 has been removed in the revised manuscript.

> *(7) Line 107: Chen et al, 2024 is listed twice in the references.*

We thank the reviewer for the careful correction. The duplicate reference of Chen et al., 2024 has been removed from the reference list in the revised manuscript.

> *(8) Line 242: remove quotation mark.*

We thank the reviewer for the careful correction. The unnecessary quotation mark at Line 242 has been removed in the revised manuscript.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *(1) Please re-analyze the data in Figures 3 and 4 using ANOVA followed by appropriate post-hoc tests (e.g., Tukey's HSD). Specifically, use a One-way ANOVA for strain comparisons in Figure 3 and a Two-way ANOVA for the learning assays in Figure 4. The interaction term (Genotype $\times$ Training) is critical for demonstrating specific learning deficits. Update the "Statistical Analysis" section and figure legends accordingly.*

We appreciate the reviewer's recommendation. We have re-analyzed the data in Figures 3 and 4 using the suggested ANOVA approaches: one-way ANOVA with Tukey's HSD for strain comparisons (Figure 3), and two-way ANOVA (including the Genotype × Training interaction term) with Sidak's post-hoc test for learning assays (Figure 4). The updated statistical methods are now described in the Statistical Analysis section and figure legends.

The revised "Statistical Analysis" section is as follows (line 512-524):

"To ensure robust statistical analysis, each experimental group included at least 100 male flies (naïve, sexually experienced, or singly reared). Internal controls were incorporated into

every experiment as recommended by Bretman et al. (2011) (Bretman et al., 2011). Normality of the mating duration data was confirmed using the Kolmogorov-Smirnov test (p>0.05). For group comparisons, two-sided Student's t-tests were applied to calculate significance levels (****p<0.0001, ***p<0.001, **p<0.01, * p<0.05), Comparisons among three or more groups were performed using one-way ANOVA with Tukey's HSD post-hoc tests. Two-factor experimental comparisons were performed using two-way ANOVA followed by Sidak's post-hoc test. while estimation statistics (Claridge-Chang and Assam, 2016) were used to visualize effect sizes, mean differences, and precision, avoiding reliance solely on null hypothesis testing. All analyses, including data plotting, were performed using GraphPad Prism software."

> *(2) To decouple motor defects from courtship deficits in w^{1118} and y^1 mutants, please use your tracking data to calculate and present a "General Locomotion" metric (e.g., average velocity or total distance traveled in the absence of a female).*

We thank the reviewer for this suggestion. We have now added Figure S4 showing basal locomotor velocity of single-housed males across all four strains. As expected, w^1118 and y^1 mutants move more slowly than wild-type controls.

These data reveal that both motor and sensory factors contribute to the observed courtship phenotypes. Reduced basal locomotion likely limits the males' ability to approach and follow females. However, this generalized hypoactivity is compounded by strain-specific sensory deficits: w^{1118} males suffer visual impairment that compromises female detection (Krstic et al., 2013), while y^1 males display altered cuticular hydrocarbons that disrupt pheromonal communication (Drapeau et al., 2006). These sensory defects impair courtship initiation and female recognition independent of locomotor capacity. Thus, the reduced CI, CL, and MD in these mutants reflect the combined effects of slower movement and courtship-specific sensory impairments, rather than motor defects alone.

We have revised the manuscript to incorporate the velocity data and clarify this interpretation (line 219-226):

"Notably, reduced basal locomotor activity in w^{1118} and y^1 mutants has been well documented in previous studies, independent of courtship behavior (Drapeau et al., 2006; Krstic et al., 2013). Consistent with these reports, our tracking data show that single-housed w^{1118} and y^1 males exhibit lower average velocity than Canton-S and Oregon-R controls (Fig. S4A). These general locomotor differences are insufficient to fully explain the observed courtship and mating timing phenotypes, indicating that additional courtship-related processes contribute to the observed behavioral differences."

> *(3) Please expand the discussion or provide a small comparative dataset contrasting DrosoMating with established tools like JAABA. Explain the specific advantages of your pipeline (e.g., cost, simplicity, focus on CL/MD) to justify its adoption over these alternatives.*

We sincerely appreciate the reviewer's insightful comments on benchmarking DrosoMating against existing tools for automated courtship behavior analysis. We fully agree that direct comparisons are essential to clarify the unique advantages and ideal use cases of DrosoMating. We have extensively revised the manuscript by adding comparative experiments, a new figure (Figure 5) and comparative table (Table 1), and expanded descriptions, as detailed below.

(1) Direct comparison with conventional tracking tools

We systematically tested Ctrax and FlyTracker on the same low-quality, high-throughput mating videos used for DrosoMating. The corresponding results are presented in Figure 5.

Ctrax showed severe segmentation instability, including over-segmentation, under-segmentation, and fragmented trajectories, and failed to stably detect two flies.

FlyTracker was able to generate background models and showed partially improved segmentation after threshold adjustment, but frequently crashed during feature computation and could not complete the end-to-end analysis pipeline.

These results confirm that tracking-based tools are vulnerable to low contrast, chamber artifacts, and prolonged male–female overlap, whereas DrosoMating bypasses pose tracking and directly detects mating states.

The revised manuscript text is as follows (line 265-342):

" DrosoMating is more compatible with low-quality mating videos than conventional tracking-based pipelines

To evaluate whether conventional fly-tracking pipelines could be used as alternative tools for extracting mating-duration metrics, we tested Ctrax and FlyTracker on representative low-quality single-chamber mating videos recorded under our standard high-throughput conditions (Fig. 5). Ctrax is designed to estimate the position and orientation of multiple walking flies while maintaining individual identities over time (Branson et al., 2009) (*https://ctrax.sourceforge.net/*), whereas FlyTracker aims to track fly pose, including position, orientation, body size, wing and leg positions, and to generate trajectory- and feature-based outputs for downstream behavior analysis (Eyjolfsdottir et al., 2014). Because these programs were not readily compatible with our full high-throughput behavioral recording setup, we first cropped the original videos and tested single-chamber videos containing one male and one female fly.

Using Ctrax, we observed that target detection was highly variable even within single-chamber videos. In representative frames, Ctrax could occasionally identify the two flies correctly (Fig. 5A). However, imperfect segmentation was frequently observed. In some frames, parts of the fly body were detected as additional targets, resulting in over-segmentation and an apparent increase in the number of detected flies (Fig. 5B, C). Conversely, when the male and female were close to each other or physically overlapped, the two animals were sometimes detected as a single target (Fig. 5D). These examples indicate that Ctrax detection was sensitive to the low contrast and overlapping fly bodies present in our mating videos. We then adjusted the Ctrax detection threshold to improve segmentation quality. Although threshold optimization improved detection in some frames, abnormal detections remained evident across randomly sampled frames (Fig. 5E). For example, some frames still showed incorrect target numbers, and a severe segmentation failure was observed in the lower-right example, where the detected objects did not correspond to two clearly separable flies. Thus, even after parameter optimization, Ctrax did not consistently maintain the expected two-target detection state in single-chamber mating videos.

The instability of Ctrax detection was also reflected in the trajectory output. In the first 500 frames, the generated trajectories were fragmented into multiple colored track segments rather than two continuous trajectories corresponding to the male and female (Fig. 5F). In addition, some trajectories extended outside the chamber boundary, indicating tracking errors and identity instability. Consistently, frame-by-frame quantification of detected target number showed that the detected object count did not remain stable at the expected value of two flies per chamber (Fig. 5G). Because Ctrax failed to maintain stable two-fly detection and continuous trajectories under these video conditions, its output could not be reliably used for downstream extraction of copulation latency or mating duration.

We next tested FlyTracker on the same type of cropped single-chamber mating videos. FlyTracker was developed to track multiple flies by estimating body position, orientation,

size, wing and leg positions, and by maintaining fly identities across video frames; it also outputs per-frame features such as velocity, facing angle, and wing-angle-related measurements for downstream behavioral analysis (Eyjolfsdottir et al., 2014). In our videos, FlyTracker was able to generate a background model, indicating that the program could recognize the overall imaging field and chamber background (Fig. 5H). However, during calibration and segmentation, the default threshold setting produced inconsistent detection results (Fig. 5I). Although some frames were segmented relatively well under the default threshold (Fig. 5J), these successful examples were not representative of the overall tracking process, and the program frequently terminated with runtime errors during tracking or downstream feature computation. To improve detection stability, we lowered the segmentation threshold. Under this adjusted setting, FlyTracker produced more complete fly masks in representative frames (Fig. 5K), and the diagnostic output showed that two flies were detected in many sampled frames (Fig. 5L). Nevertheless, detection remained unstable in some sampled frames, including frames in which zero or one fly was detected despite the expected two flies per chamber (Fig. 5L). The full pipeline ultimately failed during feature computation, producing a runtime error before complete tracking and feature outputs could be generated (Fig. 5M). Therefore, even after threshold adjustment, FlyTracker could not provide a stable end-to-end workflow for extracting mating-duration metrics from these low-quality mating videos.

This failure mode is relevant because FlyTracker depends on stable segmentation, identity maintenance, and per-frame feature extraction. In our assay videos, low contrast, chamber-edge artifacts, and prolonged male–female overlap during copulation interfered with these requirements. As a result, FlyTracker could occasionally identify the flies in individual frames, but it did not reliably complete the full analysis pipeline required for downstream behavioral quantification. This limitation is especially important for workflows such as JAABA, which use manually labeled examples to train behavior classifiers but still depend on upstream tracking-derived features. Thus, for our low-quality high-throughput mating recordings, FlyTracker-based analysis was substantially less practical than DrosoMating, which directly outputs mating-related timing metrics without requiring continuous high-fidelity two-fly pose tracking."

(2) Clarification of ideal use cases

We revised the Discussion to emphasize that DrosoMating is not intended to replace general-purpose pose or tracking tools (e.g., FlyTracker, Ctrax) that provide fine-grained behavioral data.

Instead, it offers a specialized, robust, and high-throughput workflow for scenarios requiring efficient extraction of mating timing metrics from low-quality videos, where conventional pipelines often fail.

These revisions greatly improve the clarity and positioning of DrosoMating. We thank the reviewer for this valuable suggestion.

The revised manuscript text is as follows (line 578-620):

“Comparative Evaluation with Existing Courtship Analysis Tools

A major advantage of DrosoMating is its compatibility with low-quality, high-throughput mating videos. Conventional tracking-based tools such as Ctrax and FlyTracker are powerful for trajectory- and pose-based behavioral analysis, but they generally require stable object segmentation, identity maintenance, and reliable feature extraction across frames. Ctrax was designed to estimate the positions and orientations of multiple walking flies while maintaining their identities, whereas FlyTracker aims to track detailed fly pose and generate per-frame behavioural features. Under our recording conditions, these requirements were

difficult to satisfy because the videos were low contrast and the male and female frequently overlapped during copulation.

In our tests, both tools showed limited compatibility with these videos. Even after cropping to single-chamber videos and adjusting detection parameters, Ctrax produced unstable target numbers, fragmented trajectories, and tracking errors. FlyTracker could generate a background model and occasionally segment flies successfully, but detection remained unstable and the full pipeline failed during feature computation. These issues are particularly relevant for copulation latency and mating duration analysis: during copulation, the male and female remain physically coupled for a long period, which makes identity-based tracking difficult. For these timing metrics, it is more important to robustly detect the onset and offset of the mating state than to reconstruct detailed individual trajectories. DrosoMating was purpose-built to address these specific challenges. It operates reliably on lower-quality video streams, requires no complex pre-processing or manual ROI definition, and is optimized for high-throughput multi-chamber analysis. While it does not offer the same level of pose or kinematic detail as other tools, it provides a unique solution for laboratories seeking a simple, fast, and robust pipeline to quantify core reproductive timing metrics—copulation latency (CL), courtship index (CI), and mating duration (MD)—without the overhead of more complex systems (Table 1).

Although we directly tested only Ctrax and FlyTracker, this limitation may also affect workflows that depend on upstream tracking-derived features. For example, JAABA uses tracking-derived features to train behavior classifiers, and DANCE, a recent Drosophila aggression and courtship pipeline, uses JAABA-based classifiers and lists FlyTracker and JAABA as required software. Thus, our conclusion is not that these tools are generally unsuitable for Drosophila behavior analysis, but that DrosoMating provides a more practical workflow for low-quality, high-throughput videos focused specifically on mating timing.

While DrosoMating currently prioritizes mating timing metrics over discrete behavioral classification, its modular architecture provides a foundation for future integration with behavior classifiers such as JAABA. Laboratories requiring granular behavioral elements—such as wing extension or circling—would benefit from an extended pipeline that exports per-frame kinematic features for downstream classifier training. Validating this integration represents a promising future direction to broaden the tool's utility beyond core reproductive timing assays."

We have added this clarification to the Table 1 legend to avoid overgeneralizing the limitations of Ctrax and FlyTracker (line 786-790):

"DrosoMating was designed to extract mating-related timing metrics from high-throughput videos without requiring continuous two-fly identity tracking. Ctrax and FlyTracker were tested on cropped single-chamber videos from the same recording setup. Their limitations described here refer specifically to these low-quality mating videos and should not be interpreted as general limitations of the tools."

> *(4) Complement the aggregate bar plots with behavioral ethograms or raster plots for representative individual flies. Color-code these plots for specific states (resting, following, courting, copulating) to visually demonstrate the system's temporal resolution and detection accuracy.*

We thank the reviewer for this valuable suggestion. To address this comment, we have added a new supplementary figure (Figure S3) that presents behavioral ethograms for individual flies, directly complementing the aggregate bar plots in the main text. The figure displays the full temporal progression of courtship and copulation behaviors for all wells that exhibited successful mating. As requested, behaviors are color-coded (orange: courting, red: copulating) to clearly delineate different states. These ethograms visually demonstrate the system's

ability to resolve behavioral transitions with high temporal precision, confirming the accuracy of our automated detection of courtship initiation, duration, and copulation events. This addition provides critical individual-level validation that supports the aggregate statistical results presented in the main text.

> *(5) Standardize the use of estimation statistics (DBMs). If used in Figure 3, they should also be applied to Figure 4, with appropriate statistical comparisons between groups.*

We thank the reviewer for this suggestion. We have standardized the statistical reporting format across all figure legends, with each legend explicitly stating the test used, the post hoc method (where applicable), and significance thresholds.

Our approach is as follows:

Fig. 2 and Fig.3D-I (two-group comparison, manual vs. automated scoring): DBM + Student's t-test.

Fig. 3A-C and Fig. S1B-C, E-F (multi-group comparison, 4 strains or 2 rearing conditions): One-way ANOVA + Tukey's HSD.

Fig. 4B-D and F (two-factor design, genotype × training): Two-way ANOVA + Sidak's post hoc.

We have also added a sentence to the Methods clarifying that estimation statistics (DBM) are used for single two-group contrasts, while ANOVA-based approaches are used for multi-group or multi-factor designs. The statistical methods are now consistently documented in both the Methods section and the corresponding figure legends. We hope this clarification addresses the reviewer's concern.

The revised "Statistical Analysis" section is as follows (line 512-524):

"To ensure robust statistical analysis, each experimental group included at least 100 male flies (naïve, sexually experienced, or singly reared). Internal controls were incorporated into every experiment as recommended by Bretman et al. (2011) (Bretman et al., 2011). Normality of the mating duration data was confirmed using the Kolmogorov-Smirnov test ($p>0.05$). For group comparisons, two-sided Student's t-tests were applied to calculate significance levels (****$p<0.0001$, ***$p<0.001$, **$p<0.01$, * $p<0.05$), Comparisons among three or more groups were performed using one-way ANOVA with Tukey's HSD post-hoc tests. Two-factor experimental comparisons were performed using two-way ANOVA followed by Sidak's post-hoc test. while estimation statistics (Claridge-Chang and Assam, 2016) were used to visualize effect sizes, mean differences, and precision, avoiding reliance solely on null hypothesis testing. All analyses, including data plotting, were performed using GraphPad Prism software."

> *(6) Fix inconsistent labeling (e.g., "MB-247" vs. "MB247") and redundant axis labels.*

Thank you for pointing this out. We have now standardized all labels to "MB247-GAL4" throughout the text and figures.

We have also removed redundant axis labels from multi-panel figures. And redundant DBM and metric labels have been streamlined in Figures 2 and 4. All statistical tests and metric definitions are now fully described in the corresponding figure legends rather than being repeated on each sub-panel.

> *(7) Significantly increase the size of data panels to make individual data points and error bars legible.*

Thank you for this suggestion. We have standardized the figure formatting and simplified the panels by removing redundant axis labels and consolidating descriptive details into the

figure legends. We believe the current panel sizes, combined with these clarifications, provide sufficient legibility for both data points and error bars in the final high-resolution PDF.

> *Minor Corrections:*
>
> *(1) Abstract: Rephrase "lack of certain timing-related behavioral repertoires" (Line 108) to "lack of precise quantification for temporal parameters of post-copulatory behavior."*

Thank you for the suggested rephrasing. We have updated the sentence accordingly.

> *(2) Define the physical parameter "s" (e.g., is it a pixel threshold?) to ensure reproducibility.*

Thank you for this important suggestion. We have now explicitly defined the parameter s in the Methods section. Briefly, s is the grayscale intensity threshold (0–255, 8-bit) used for binary segmentation of flies from the background. It is automatically calculated as the maximum grayscale value of three user-selected reference flies plus an offset of 28, and can be manually adjusted to accommodate varying illumination conditions.

The revised manuscript text is as follows (line 499-503):

“Note on the s value: The parameter s represents the grayscale intensity threshold (range: 0–255 for 8-bit images) used for binary segmentation of flies from the background. It is automatically calculated as the maximum grayscale value at the three reference fly positions plus an offset of 28, and can be manually adjusted to accommodate varying illumination conditions.”

> *(3) Figure 1 Legend: Clarify the "clockwise selection" of pillars and their relation to well numbering.*

Thank you for this suggestion. We have revised the Figure 1 legend to clarify that:

(1) The four pillars are selected in clockwise order starting from the top-left corner to define the chamber corners for perspective transformation (homography), which corrects for camera angle and standardizes the field of view.

(2) Well numbering is independent of pillar selection order. After automated perspective correction, wells are numbered sequentially in a left-to-right, top-to-bottom order (1–36) based on the standardized chamber layout.

The updated Figure 1 legend now reads:

"Columns (pillars) should be selected in clockwise order starting from the top-left corner to define the chamber boundaries for perspective transformation (Fig. 1A, lower). Well numbering (1–36) follows a left-to-right, top-to-bottom sequence after perspective correction and is independent of pillar selection order."

> *(4) Add the citation for Eastwood and Burnet (1977) regarding Courtship Latency.*

Thank you for pointing this out. We have added the citation Eastwood and Burnet (1977) to the Introduction where Courtship Latency is first defined

> *(5) Select one term ("Mating Duration" or "Copulation Duration") and use it consistently throughout the text and figures.*

Thank you for this suggestion. We have now standardized the terminology throughout the manuscript and figures. "Mating Duration" (MD) is used consistently in all instances where

"Copulation Duration" previously appeared. The abbreviation MD has been retained for consistency with existing figure labels

References

Branson, K., Robie, A.A., Bender, J., Perona, P., & Dickinson, M.H. (2009). High-throughput ethomics in large groups of Drosophila. Nature Methods, 6(6), 451–458.

Claridge-Chang, A., & Assam, P.N. (2016). Estimation statistics should replace significance testing. Nature Methods, 13(2), 108–109.

Drapeau, M.D., Cyran, S.A., Viering, M.M., Geyer, P.K., & Long, A.D. (2006). A cis-regulatory sequence within the yellow locus of Drosophila melanogaster required for normal male mating success. Genetics, 172(2), 1009–1030.

Eastwood, L., & Burnet, B. (1977). Courtship latency in male Drosophila melanogaster. Behavior Genetics, 7(3), 359–372.

Eyjolfsdottir, E., Branson, S., Burgos-Artizzu, X.P., Hoopfer, E.D., Schor, J., Anderson, D.J., & Perona, P. (2014). Detecting social actions of fruit flies. Lecture Notes in Computer Science, 8692, 772–787.

Gil-Martí, B., Barredo, C.G., Pina-Flores, S., Poza-Rodriguez, A., Treves, G., Rodriguez-Navas, C., Camacho, L., Pérez-Serna, A., Jimenez, I., Brazales, L., Fernandez, J., & Martin, F.A. (2023). A simplified courtship conditioning protocol to test learning and memory in Drosophila. STAR Protocols, 4(1), 101572.

Greenspan, R.J., & Ferveur, J.F. (2000). Courtship in Drosophila. Annual Review of Genetics, 34, 205–232.

Hall, J.C. (1994). The mating of a fly. Science, 264(5163), 1702–1714.

Kabra, M., Robie, A.A., Rivera-Alba, M., Branson, S., & Branson, K. (2013). JAABA: interactive machine learning for automatic annotation of animal behavior. Nature Methods, 10(1), 64–67.

Kitamoto, T. (2001). Conditional modification of behavior in Drosophila by targeted expression of a temperature-sensitive shibire allele in defined neurons. Journal of Neurobiology, 47(2), 81–92.

Krstic, D., Boll, W., & Noll, M. (2013). Influence of the White locus on the courtship behavior of Drosophila males. PLoS ONE, 8(9), e77904.

Levin, L.R., Han, P.L., Hwang, P.M., Feinstein, P.G., Davis, R.L., & Reed, R.R. (1992). The Drosophila learning and memory gene rutabaga encodes a Ca2+/calmodulin-responsive adenylyl cyclase. Cell, 68(3), 479–489.

Pavlou, H.J., & Goodwin, S.F. (2013). Courtship behavior in Drosophila melanogaster: towards a 'courtship connectome'. Current Opinion in Neurobiology, 23(1), 76–83.

Yapici, N., Kim, Y.J., Ribeiro, C., & Dickson, B.J. (2008). A receptor that mediates the post-mating switch in Drosophila reproductive behaviour. Nature, 451(7176), 33–37.

https://doi.org/10.7554/eLife.109742.2.sa0