

DANCE: An open-source analysis pipeline and low-cost hardware to quantify aggression and courtship in *Drosophila*

Reviewed Preprint

v1 • March 25, 2025

Not revised

R Sai Prathap Yadav, Paulami Dey, Faizah Ansari, Tanvi Kottat, P Pallavi Prabhu, Manohar Vasam, Shrinivas Ayyangar, Swathi Bhaskar S, Krishnananda Prabhu, Monalisa Ghosh, Pavan Agrawal 

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India • Department of Biochemistry, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

eLife Assessment

This study presents a **valuable** open-source and cost-effective method for automating the quantification of male aggression and courtship in *Drosophila melanogaster*. The work as presented provides **solid** evidence that the use of the behavioral setup that the authors designed - using readily available laboratory equipment and standardised high-performing classifiers they developed using existing software packages - accurately and reliably characterises social behavior in *Drosophila*. The work will be of interest to *Drosophila* neurobiologists and particularly to those working on male social behaviors.

<https://doi.org/10.7554/eLife.105465.1.sa3>

Abstract

Quantifying animal behaviors is pivotal for identifying the underlying neuronal and genetic mechanisms. Computational approaches have enabled automated analysis of complex behaviors such as aggression and courtship in *Drosophila*. However, existing approaches rely on rigid, rule-based algorithms and expensive hardware, limiting sensitivity to behavioral variations and accessibility. Here, we describe the DANCE (*Drosophila* Aggression and Courtship Evaluator), a low-cost, open-source platform combining machine learning-based classifiers and inexpensive hardware to quantify aggression and courtship. DANCE consists of six novel behavioral classifiers trained using a supervised machine learning algorithm. DANCE classifiers outperform existing rule-based algorithms by capturing dynamic behavioral variations. DANCE hardware is constructed using repurposed medicine blister packs and acrylic sheets, with recordings performed using smartphones, making it affordable and accessible. Benchmarking demonstrated that DANCE hardware performs comparably to sophisticated, high-cost setups. We validated DANCE in diverse contexts, including social isolation versus enrichment that modulate aggression and courtship, RNAi-mediated down-regulation of neuropeptide Dsk, and optogenetic silencing of dopaminergic neurons which

promoted aggression. DANCE provides a cost-effective and portable solution for studying *Drosophila* behaviors in resource-limited settings or closer to natural habitats. Its accessibility and robust performance democratizes behavioral neuroscience, enabling rapid screening of genes and neuronal circuits underlying complex social behaviors.

Introduction

Detailed and accurate annotation and analysis of complex behaviors is necessary for understanding the underlying neural and molecular mechanisms. Fruit fly *Drosophila melanogaster* is one of the most accessible and well-studied model organisms for identifying neuronal and molecular underpinnings of behavior. Multiple large-scale screens have been conducted using *Drosophila*, to study complex social behaviors such as aggression and courtship (Asahina, 2017; Greenspan and Ferveur, 2000; Hall, 2002; Kravitz and Fernandez, 2015) to identify underlying neural circuitry (Agrawal et al., 2020; Asahina et al., 2014; Davis et al., 2018; Hoopfer et al., 2015; Yadav et al., 2024) and genes involved (Agrawal et al., 2020; Benzer, 1967; Gill, 1963; Hall, 1978; Ishii et al., 2022; Wang et al., 2008). These behaviors exhibit stereotyped patterns. For instance, aggression involves chasing, fencing (Jacobs, 1960), wing threats, boxing (Dow and Schilcher, 1975), lunging, and tussling (Hoffmann, 1987a, 1987b). Similarly, courtship consists of multiple stereotyped behaviors exhibited by the male fly, such as orienting, circling, and following the female (Cook and Cook, 1975; Markow, 1987; O'Dell, 2003). To stimulate the female to be more receptive, the male produces a species-specific song by vibrating and extending its wing (Bennet-Clark and Ewing, 1969; Swain and von Philipsborn, 2021). The male then attempts copulation by curling its abdomen and finally mounts the female for copulation (Bastock and Manning, 1955; Spieth, 1974).

Manual analysis by trained observers is considered the gold standard in behavioral analysis, but it is a time-consuming process unsuitable for large-scale screens (Gomez- Marin et al., 2014; Robie et al., 2017a). ‘Computational ethology’ (Anderson and Perona, 2014; Datta et al., 2019) helps address this challenge by automating behavioral annotation by leveraging advances in computer vision and machine learning (Robie et al., 2017b). This enables high-throughput behavioral screening to identify responsible genes and circuits. A typical computational ethology workflow involves recording animal behaviors and tracking their positions along with body movements. This is followed by the analysis and classification of the observed behaviors from hundreds to thousands of video frames capturing behavioral instances. Several software such as Ctrax, Caltech FlyTracker, Deep Lab Cut etc. (Branson et al., 2009; Eyjolfsson et al., 2014; Mathis et al., 2018) are used widely for tracking behaviors in *Drosophila*. Each comes with their strengths and weaknesses.

For instance, Ctrax (Branson et al., 2009) helps track fly body position and movements but leads to frequent identity switches among tracked flies, which is not the case while using FlyTracker (Eyjolfsson et al., 2014). Effectiveness of various machine learning pipelines is eventually measured by comparing their output to human annotation, called— ‘ground- truthing’. Rule-based algorithms such as CADABRA (Dankert et al., 2009) is used to quantify aggression, but it can lead to mis-scoring and identity switches, as revealed by ground-truthing (Simon and Heberlein, 2020), which needs to be corrected in a semi- automated manner (Kim et al., 2018). MateBook (Ribeiro et al., 2018) is another rule-based algorithm to quantify courtship; however, similar to CADABRA, it tends to miss true-positive events, leading to significant mis-scoring of behaviors.

Janelia Automatic Animal Behavior Annotator (JAABA) (Kabra et al., 2013) addresses the challenges of rigid rule-based approaches by employing a supervised learning approach. In the JAABA pipeline, user-labeled data is utilized for training to encompass the dynamic variations in behaviors, allowing it to predict behaviors based on learning from input data.

Recently, a few JAABA-based behavioral classifiers have been developed for measuring aggression (Chowdhury et al., 2021 [↗](#); Leng et al., 2020 [↗](#)) and courtship (Gil-Martí et al., 2023 [↗](#); Pantalia et al., 2023 [↗](#)). However, these studies either do not make these classifiers publicly available (Gil-Martí et al., 2023 [↗](#); Leng et al., 2020 [↗](#); Pantalia et al., 2023 [↗](#)) or require specialized hardware such as 3D printed parts (Chowdhury et al., 2021 [↗](#); Gil-Martí et al., 2023 [↗](#)) and expensive machine vision cameras (Chowdhury et al., 2021 [↗](#); Leng et al., 2020 [↗](#)), making them inaccessible for resource-limited settings and limiting their wider adoption.

Here we describe DANCE (*Drosophila* Aggression and Courtship Evaluator), an open-source, user-friendly analysis and hardware pipeline to simplify and automate the process of robustly quantifying aggression and courtship behaviors. DANCE has two components: 1. A set of robust, machine vision-based behavioral classifiers developed using JAABA to quantify aggression and courtship; and 2. An inexpensive hardware utilizing off-the-shelf components to construct behavioral arenas and smartphone cameras to record behaviors. DANCE classifiers outperform previous methods (Dankert et al., 2009 [↗](#); Ribeiro et al., 2018 [↗](#)), improving accuracy and reliability, and DANCE hardware eliminates the need for complex, specialized behavioral arenas and expensive machine vision cameras to assess fly behaviors. Therefore, the DANCE assay can be useful for wider adoption, including in resource-strapped settings. DANCE provides a platform for rapid behavioral screening using the *Drosophila* model for discovering mechanisms underlying complex social behaviors and neurological disorders.

Results

DANCE assay analysis pipeline

To overcome the challenge of time-consuming manual behavioral annotation or resource-intensive, complex hardware, we developed an automated, high-throughput quantification pipeline—DANCE and trained new behavioral classifiers using an existing machine learning algorithm—JAABA (Kabra et al., 2013 [↗](#)), to robustly quantify aggression and courtship in *Drosophila* (Figure 1 [↗](#)). We built a novel hardware platform consisting of repurposed transparent medicine tablet foils/blister packs, acrylic sheets, and paper tape to easily record aggression and courtship behaviors. To record these behaviors we used Android smartphone cameras and an electronic tablet or smartphone served as a backlight illumination source (Figure 1A [↗](#), Materials and Methods). We quantify these behaviors using our ‘DANCE classifiers’. Unlike existing set-ups that cost approximately USD 3500, ‘DANCE hardware’ is fabricated using off-the-shelf components and costs less than USD 30 cents (Supplementary file 1). To benchmark the performance of the DANCE classifiers, we used pre-existing setups and rule-based methods for quantifying courtship and aggression and compared their performance with DANCE classifiers (Figure 1B-C [↗](#)).

To train DANCE classifiers using JAABA (Kabra et al., 2013 [↗](#)), for aggressive lunges, we used a pre-existing setup described in (Dankert et al., 2009 [↗](#)) modified from (Dierick, 2007 [↗](#)) (Figure 1A [↗](#); Figure 1—figure supplement 1 [↗](#)); and for courtship behaviors, we used a pre-existing set up described in (Koemans et al., 2017 [↗](#)) (Figure 1A [↗](#); Figure 1—figure supplement 2 [↗](#)). We tracked position, motion, and interactions of pairs of flies across video frames using the Caltech FlyTracker (Eyjolfsson et al., 2014 [↗](#)). To avoid data leak, we randomly divided acquired videos into two categories: “training videos” and “test videos” to train and evaluate DANCE classifiers. These test videos were also manually ‘ground-truthed’ frame by frame, which is considered the gold standard for behavioral annotation (Figure 1B, D-E [↗](#)). We benchmarked the performance of DANCE classifiers using existing rule-based algorithms, CADABRA (Dankert et al., 2009 [↗](#)) and MateBook (Ribeiro et al., 2018 [↗](#)); and an existing JAABA aggression classifier (Chowdhury et al., 2021 [↗](#)). We then compared their output against manual ground-truth. Figure 1 D-E [↗](#) show examples of courtship and aggression video annotation, suggesting that DANCE classifiers

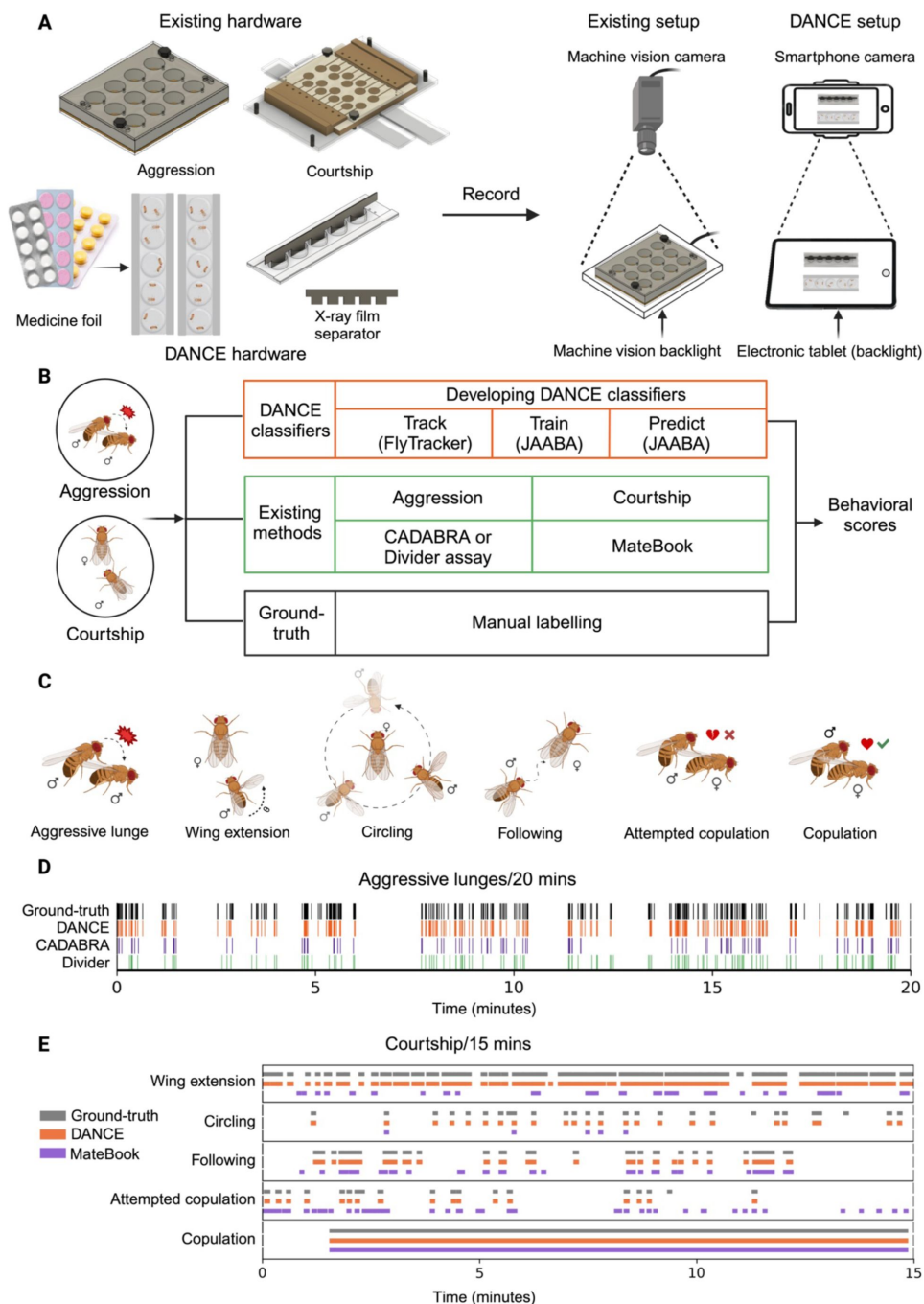


Figure 1.

DANCE assay overcomes the limitation of existing methods for quantification of aggression and courtship behaviors.

(A) Existing hardware for acquisition of aggression and courtship behavior using machine vision cameras and the simplified DANCE hardware. (B) Steps for developing DANCE classifiers and benchmarking with existing methods and manual ground-truthing to produce behavioral scores. (C) Various behavioral classifiers developed to quantify male aggression (lunge) and male courtship (wing extension, circling, following, attempted copulation, and copulation). (D) Raster plots comparing performance of ground-truth vs. DANCE classifier vs. CADABRA vs. Divider assay for aggression in a representative video. (E) Raster plots for various courtship behaviors comparing performance of ground-truth vs. DANCE classifiers vs. MateBook from representative videos. Created in *BioRender*.

outperform existing methods and are comparable with manual ground-truth. Subsequent sections describe the quantitative analysis of individual DANCE classifiers and benchmarking of DANCE hardware.

DANCE lunge classifier to quantify aggressive behavior

Aggression is an innate, complex behavior, and *Drosophila* males exhibit several stereotyped behavioral patterns during aggressive encounters, with lunging used widely as a measure of overall aggression in males (Agrawal et al., 2020 [↗](#); Asahina et al., 2014 [↗](#); Chiu et al., 2021 [↗](#); Chowdhury et al., 2021 [↗](#); Davis et al., 2014 [↗](#); Dierick, 2007 [↗](#); Hoffmann, 1987a [↗](#); Hoopfer et al., 2015 [↗](#); Hoyer et al., 2008 [↗](#); Jung et al., 2020 [↗](#); Nilsen et al., 2004 [↗](#); Watanabe et al., 2017 [↗](#); Yadav et al., 2024 [↗](#)). A lunge is defined as a male fly raising its front legs and hitting down on the other fly. We developed a new classifier using JAABA (Kabra et al., 2013 [↗](#)) to robustly quantify aggressive lunges in *Drosophila*, hereafter referred to as the ‘DANCE lunge classifier.’ We quantified lunges using our classifier from 20-minute-long videos and compared the output with manual ground-truthing and existing methods— CADABRA (Dankert et al., 2009 [↗](#)) and ‘Divider assay classifier’ (Chowdhury et al., 2021 [↗](#)). CADABRA tends to miss several true positive lunges, likely due to its rigid, rule-based framework, which cannot adapt to the dynamic variations in behavior. To address this, ‘Divider assay classifier’ was developed based on JAABA, but the setup requires use of 3- D printed rectangular arena and dividers as well as expensive machine vision cameras (Chowdhury et al., 2021 [↗](#)).

Figure 2A [↗](#) shows lunge scores from 40 different videos using ground-truth, DANCE lunge classifier, CADABRA, and Divider assay classifier. While the ground-truth and DANCE classifier’s output are comparable, CADABRA and the Divider assay classifier underscored lunges across videos. We ground truthed the DANCE lunge classifier against 40 ‘test videos’ (**Figure 2A** [↗](#), Materials and Methods). To address potential observer bias during ground-truthing, we compared the annotations between two separate evaluators for the DANCE lunge classifier and found no significant differences (**Figure 2— figure supplement 1A** [↗](#)). Since aggressive lunges have a large dynamic range, we benchmarked our classifier across a range of aggressive behaviors, and subdivided the ground truthed videos into four categories: 1. low aggressive, 0–70 lunges (**Figure 2B** [↗](#)); 2. moderately aggressive, 71–160 lunges (**Figure 2C** [↗](#)); 3. highly aggressive, 161–300 lunges (**Figure 2D** [↗](#)); and 4. hyper- aggressive, >300 lunges (**Figure 2E** [↗](#)). We found that DANCE lunge classifier and manual ground-truth performance was comparable across all four categories (**Figure 2 B-E** [↗](#)). In contrast, the CADABRA and Divider assay classifier showed significant underscoring from the manual ground-truth (**Figure 2B-E** [↗](#)), however, a strong correlation was observed between our DANCE lunge classifier and ground-truth (**Figure 2F** [↗](#), Supplementary file 2), whereas, CADABRA and the Divider assay classifier showed only a weaker correlation with the ground truth (**Figure 2G-H** [↗](#), Supplementary file 2). We reasoned that CADABRA’s lower performance is most likely due to the rigid rules used to define a lunge (Dankert et al., 2009 [↗](#)) (Materials and Methods); and the Divider assay classifier’s trained in a rectangular arena did not perform as well in a circular arena since JAABA classifiers utilize features related to arena geometry.

To further evaluate the performance of our DANCE lunge classifier, we used three metrics— precision, recall, and F1 score and compared the output of the classifier and manual ground-truth across 40 test videos. We calculated the counts of true positives, false positives, and false negatives to derive precision, which depicts the ratio of true positives among overall predictions from the classifier; recall, which depicts the ratio of true positives among the ground-truth scores; and F1 score, which is a harmonic mean of precision and recall (**Figure 2I** [↗](#)) (Materials and Methods). Our DANCE lunge classifier had a precision of 76.80%, a recall rate of 71.53% and an overall F1 score of 73.60% (**Figure 2I** [↗](#)), which is higher than the CADABRA and the Divider assay suggesting that the DANCE lunge classifier outperforms these existing methods. Together, our analysis suggests that the DANCE lunge classifier performs with high precision and quantifies lunge numbers robustly over a broad range of fighting intensities.

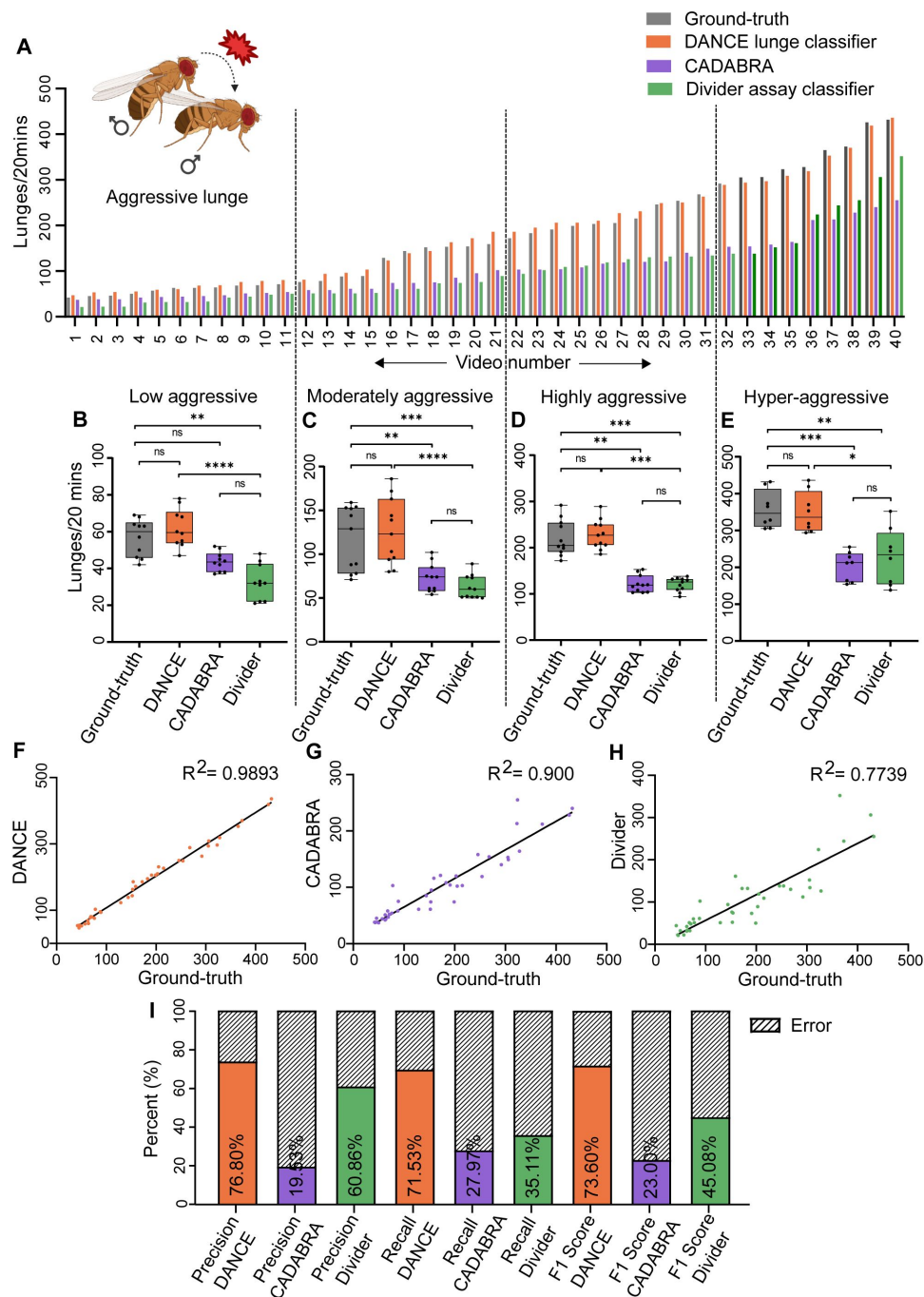


Figure 2.

DANCE lunge classifier outperforms existing methods of quantifying male aggression.

(A) Lunge scores from 20 minute long videos from ground-truth (grey), DANCE lunge classifier (orange), CADABRA (purple) and Divider assay classifier (green). (B-E) Comparison of lunge scores from different levels of aggression. Male flies showing lunges ranging from 0 to 500 were compared manually, with a new DANCE lunge classifier, CADABRA, and Divider assay classifier. (n=10; **P<0.0017, ****P<0.0001). (C) 71–160 lunges, 'moderately aggressive', (n=11; **P<0.0102, ****P<0.0002, ****P<0.0001). (D) 161–300 lunges, 'highly aggressive', (n=11; **P<0.0057, ****P<0.0002, ****P<0.0004). and (E) >300 lunges, 'hyper-aggressive', (n=8; *P<0.0402, **P<0.0029, ****P<0.0006; Friedman's ANOVA with Dunn's test) (F) Regression analysis of DANCE 'lunge classifier' vs. manual scores ($R^2=0.9893$, n=40). (G) CADABRA vs. DANCE lunge classifier ($R^2=0.9$, n=40). (H) Divider assay lunge classifier vs. manual scores ($R^2=0.7739$, n=40). (I) F1 score, precision, and recall between DANCE lunge classifier, CADABRA, and Divider assay classifier.

DANCE classifiers to quantify courtship behaviors in *Drosophila*

The first report of *Drosophila* courtship behavior, identified stereotypic behaviors such as wing ‘scissor-like’ movements, males ‘swaying around the female,’ licking, tapping, and mounting (Sturtevant, 1915 [↗](#)). By the 2000s, studies uncovered genes and neural circuits involved in courtship (Dickson, 2008 [↗](#); Pavlou and Goodwin, 2013 [↗](#)). Automated analysis techniques for courtship exist but face barriers to adoption due to expensive hardware, reliance on 3D printing, or lack of publicly available analysis code or classifiers (Gil-Martí et al., 2023 [↗](#); Koemans et al., 2017 [↗](#); Leng et al., 2020 [↗](#); Reza et al., 2013 [↗](#)).

MateBook is a recent rule-based algorithm that was developed to automate quantifying various courtship behaviors in *Drosophila* males (Ribeiro et al., 2018 [↗](#)). It relies on predefined mathematical rules based on CADABRA (Dankert et al., 2009 [↗](#)). To resolve ambiguities in the two overlapping flies during estimating their trajectories from the video recordings, it assigns the identities by relying on the distinct body sizes of the male and female fly, as the females are larger than males. However, decapitated virgin females are of similar body size as males, and are often used in courtship assays to assess the male attractiveness to other factors such as pheromones, independent of the female behavior (Cook and Cook, 1975 [↗](#); Spieth, 1966 [↗](#)). In such cases, rigid rule-based approach can introduce both false positives and false negatives since the condition about differential body size is not met.

To address the inherent inflexibility of rule-based algorithms, we trained and validated five new courtship classifiers using JAABA (Kabra et al., 2013 [↗](#)) to provide a detailed and robust analysis of multiple steps of courtship ritual (Sokolowski, 2001 [↗](#)) including— wing extension, following, circling, attempted copulation, and copulation. We evaluated the performance of our DANCE classifiers by comparing the output to ground-truth and MateBook (Ribeiro et al., 2018 [↗](#)). Criteria used to develop various classifiers are described in Materials and Methods. To compare the output, we calculated a behavior index, since courtship behaviors, unlike aggressive lunges, are of varied duration. To address potential observer bias during ground-truthing, we compared the annotations between two separate evaluators for the DANCE courtship classifiers and found no significant differences, suggesting robustness of our classifiers (**Figure 2— figure supplement 1B-F** [↗](#)).

Wing Extension

During unilateral wing extension, the male vibrates its wing at a specific frequency to produce a species-specific courtship song to attract the female (Shorey, 1962 [↗](#); Spieth, 1952 [↗](#)). **Figure 3A** [↗](#) shows wing extension index from the ground-truth (grey) and the ‘DANCE wing extension classifier’ (orange), which are comparable across 15 videos from males courting decapitated virgin females, however, MateBook’s output (purple) seems to underscore these events in most videos (**Figure 3A** [↗](#), Friedman’s ANOVA with Dunn’s test; ground-truth vs. MateBook, $p=0.0020$, $n=15$; ground-truth vs. DANCE classifier, $p>0.9999$) (**Figure 3A** [↗](#)). We also evaluated the overall wing extension index from all 15 videos and found ground-truth scores and the DANCE wing extension classifier matched well, but MateBook significantly underscoring wing extension (**Figure 3B** [↗](#)). We found a strong correlation between the DANCE wing extension classifier index to ground-truth (**Figure 3C** [↗](#)) but a weaker correlation with MateBook (**Figure 3D** [↗](#)). To further evaluate our classifier’s performance, we calculated precision, recall, and F1 scores, as described earlier. The DANCE wing extension classifier had a precision of 92.19%, a recall rate of 98.09%, and an overall F1 score of 95.05% when compared to the ground-truth which is higher than the MateBook (**Figure 3E** [↗](#)).

Quantification of the videos containing mated females identified a robust correlation between our classifier’s output and the ground-truth, whereas a general trend of underscoring was observed for MateBook (**Figure 3— figure supplement 1** [↗](#)). Together, this suggests that the DANCE wing

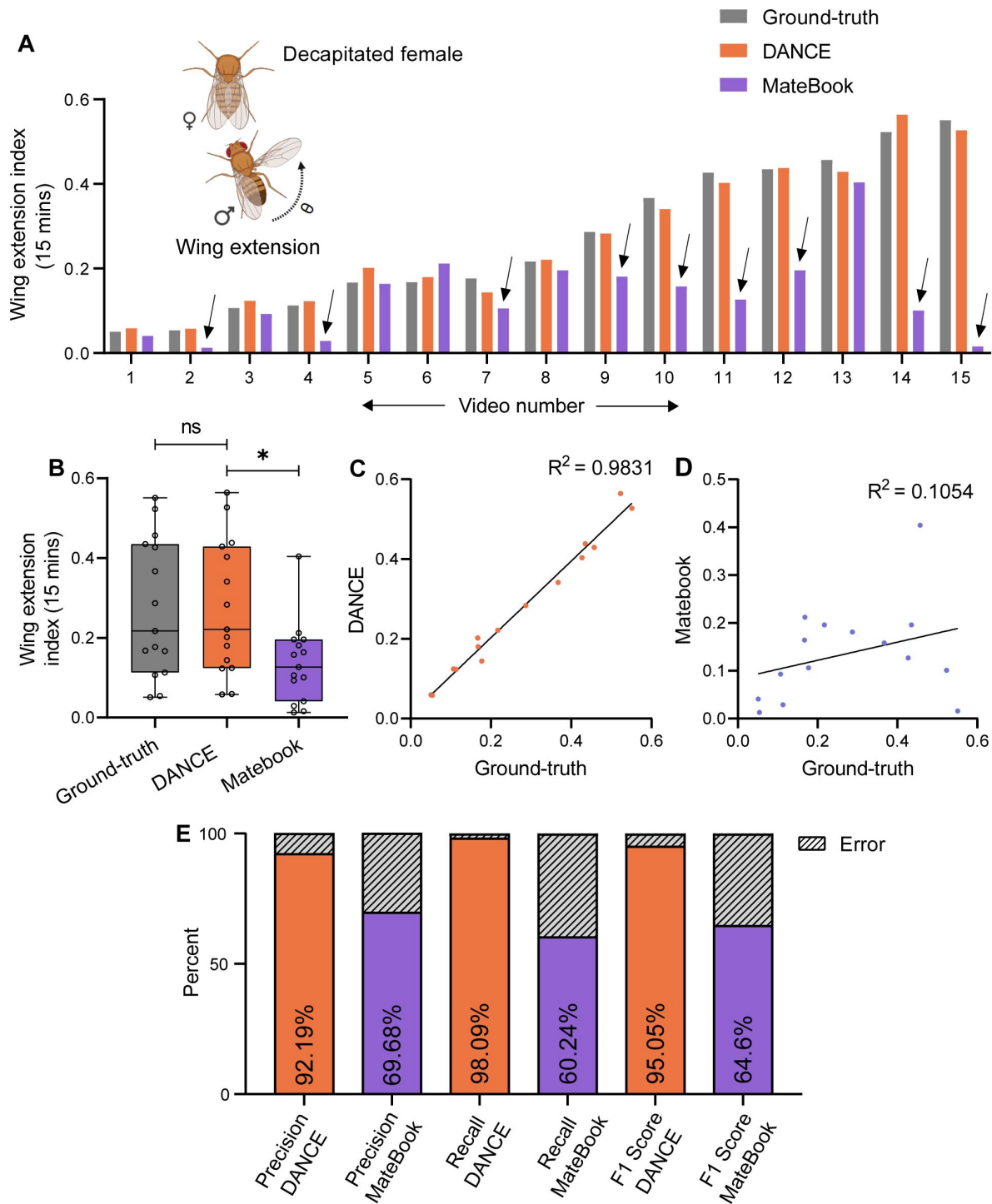


Figure 3.

DANCE wing extension classifier outperforms existing methods of quantification.

(A) Wing extension index of males from 15 minutes long videos from the ground-truth (grey), DANCE wing extension classifier (orange) and MateBook (purple), against decapitated virgin females. MateBook underscored wing extension across multiple videos (black arrows). (B) Comparison of ground-truth vs. DANCE vs. MateBook wing extension classifier (Kruskal-Wallis ANOVA with Dunn's test, ns, $p > 0.9999$, $*p = 0.0436$; $n = 15$). (C) Regression analysis of the DANCE wing extension classifier vs. ground-truth ($R^2 = 0.9831$, $n = 15$). (D) MateBook vs. ground-truth ($R^2 = 0.1054$, $n = 15$). (E) F1 score, precision, and recall of DANCE wing extension classifier and MateBook against ground-truth scores.

extension classifier is robust for both decapitated and mated females and performs with high precision.

Attempted copulation and copulation

Copulation in male flies lasts for about 15–25 minutes, and its duration is primarily determined by the male (MacBean and Parsons, 1967 [↗](#)). Interrupted mating experiments show that the sperms are transferred several minutes after copulation begins (Fowler, 1973 [↗](#); Tompkins et al., 1980 [↗](#)). Thus this differentiates the mounting behavior into two— successful copulation and unsuccessful attempted-copulation. We developed the ‘DANCE attempted copulation’ with a threshold of 0.33 to <45 seconds and the ‘DANCE copulation’ classifiers with a threshold of >45 seconds, based on existing definitions (Ribeiro et al., 2018 [↗](#)), but with modifications, since MateBook only quantifies ‘copulation’ and not ‘attempted-copulation’ (Materials and Methods). For ground-truthing, we combined both the dataset of mated and decapitated virgin females, because we observed that MateBook consistently overscored in both datasets. (**Figure 4A–B** [↗](#)). MateBook’s rule-based approach led to significant overscoring with reduced thresholds for attempted copulations, especially with decapitated virgins, due to both increased false positives and identity switches. As expected, the DANCE classifier matched well with the ground-truth scores for each individual video (Figure 4A— Friedman’s ANOVA with Dunn’s test; $p > 0.9999$, $n = 32$), as well as the overall attempted copulation index (**Figure 4B** [↗](#)). We found a strong correlation between ground-truth and the DANCE attempted copulation classifier index (**Figure 4C** [↗](#)) but a weaker correlation with MateBook (**Figure 4D** [↗](#)). The DANCE classifier performed robustly with a precision rate of 82.55%, a recall rate of 89.24%, and a F1 score of 85.77% as compared to MateBook (**Figure 4E** [↗](#)). Together, this suggests that the DANCE attempted copulation outperforms existing rule based framework and is able to capture dynamic variations in behaviors robustly.

Since the male readily copulates with the virgin females for several minutes, we used videos where males were paired with virgin females to train the ‘DANCE copulation classifier’. Our classifier matched well with the ground truth in the number of bouts, with a 100% precision, recall and F1 score (**Figure 4— figure supplement 1** [↗](#)). The rule-based approach of MateBook also works well with copulation, which MateBook defines as ‘occlusion’ persisting for a long time (Materials and Methods). However sometimes, misaligned videos can cause ‘arena detection’ errors, causing missed events (**Figure 4— figure supplement 1A** [↗](#), video#18) or false positives (**Figure 4— figure supplement 1** [↗](#), video#13) leading to lower precision, recall and F1 scores.

Circling

The male circles around the female, which is a unique male behavior during courtship. Studies have shown that a refractory female gets restimulated by the circling behavior of the males (Kessler, 1962 [↗](#)). In addition, differences in the frequency of circling behavior and other courtship rituals is used to maintain reproductive isolation among different species (Brown, 1965 [↗](#)).

To evaluate the ‘DANCE circling classifier’, we used both the decapitated virgin female (**Figure 5** [↗](#)) and mated female dataset (**Figure 5— Figure Supplement 1** [↗](#)). The circling index from the ground-truth and the DANCE classifier output are comparable; however, MateBook underscoring these events (**Figure 5A** [↗](#)), most likely due to identity switches. DANCE circling classifier matched well from the ground-truth for individual videos (**Figure 5A** [↗](#), Friedman’s ANOVA with Dunn’s test; $p < 0.2049$, $n = 12$). Overall circling index from all 12 videos from the ground-truth scores and the DANCE circling classifier were comparable (**Figure 5B** [↗](#)), whereas MateBook significantly underscores circling in individual videos (**Figure 5A** [↗](#), Friedman’s ANOVA with Dunn’s test; $***p < 0.0001$, $n = 12$). We found a strong correlation between the DANCE circling classifier index to ground-truth (**Figure 5C** [↗](#)) but a weaker correlation with MateBook (**Figure 5D** [↗](#)). The DANCE circling classifier has a precision rate and a recall rate of 98.04% and 92.09%, with a F1 score of 94.97% (**Figure 5E** [↗](#)), suggesting robustness of the classifier.

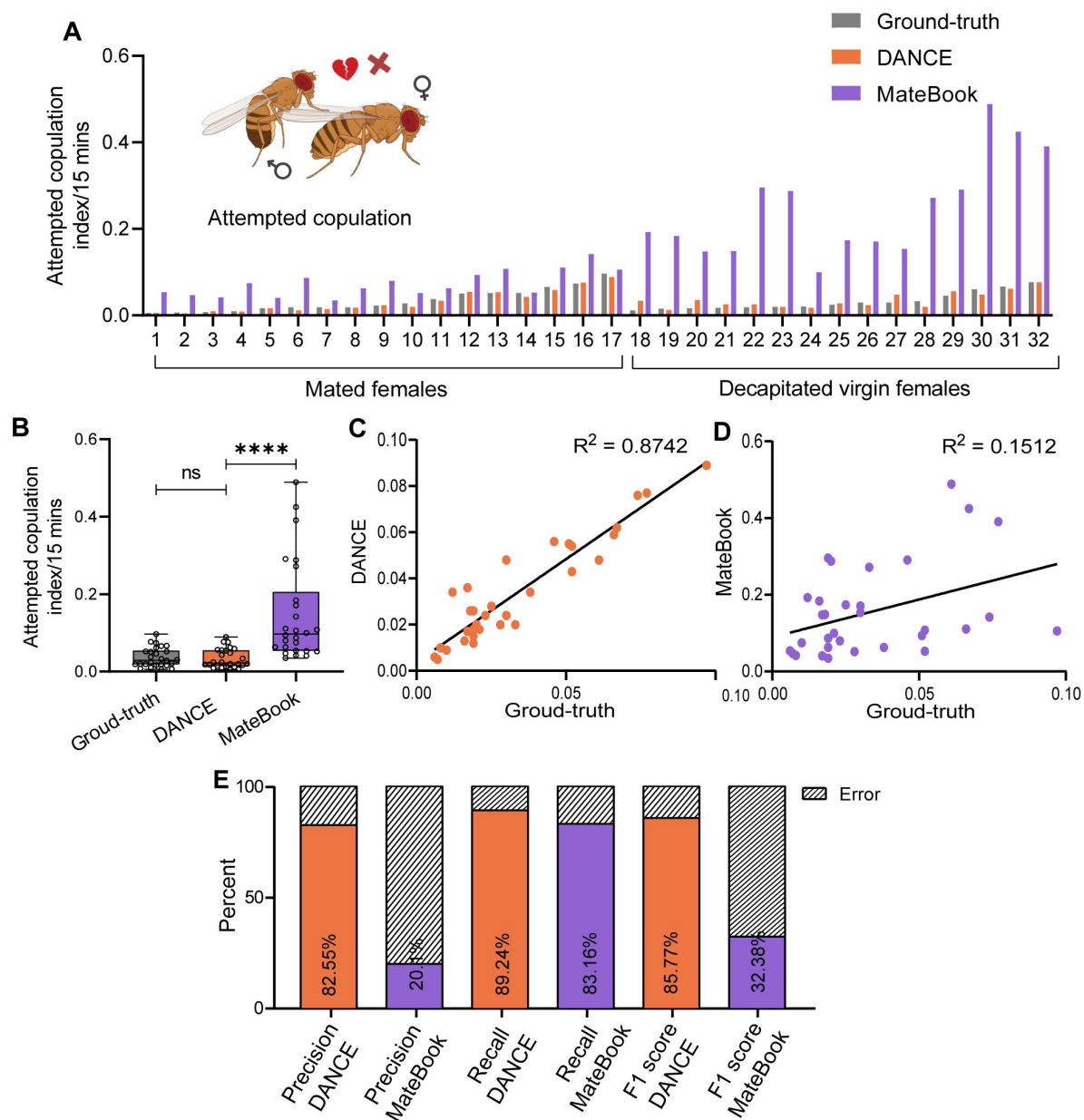


Fig. 4.

DANCE attempted-copulation classifier outperforms existing methods of quantification.

(A) Attempted copulation index of males from 15 minute long videos from the ground-truth (grey), 'DANCE attempted copulation classifier' (orange) and MateBook (purple) against both mated and decapitated females. (B) Comparison of ground-truth vs. DANCE attempted-copulation classifier vs. MateBook (Kruskal-Wallis ANOVA with Dunn's test, ns, $p > 0.9999$; **** $p < 0.0001$, $n = 32$). (C) Regression analysis of the attempted-copulation classifier vs. ground-truth ($R^2 = 0.8742$, $n = 32$). (D) Regression analysis of MateBook vs. ground-truth ($R^2 = 0.1512$, $n = 32$). (E) F1 score, precision, and recall of DANCE and MateBook attempted-copulation classifiers against ground-truth scores.

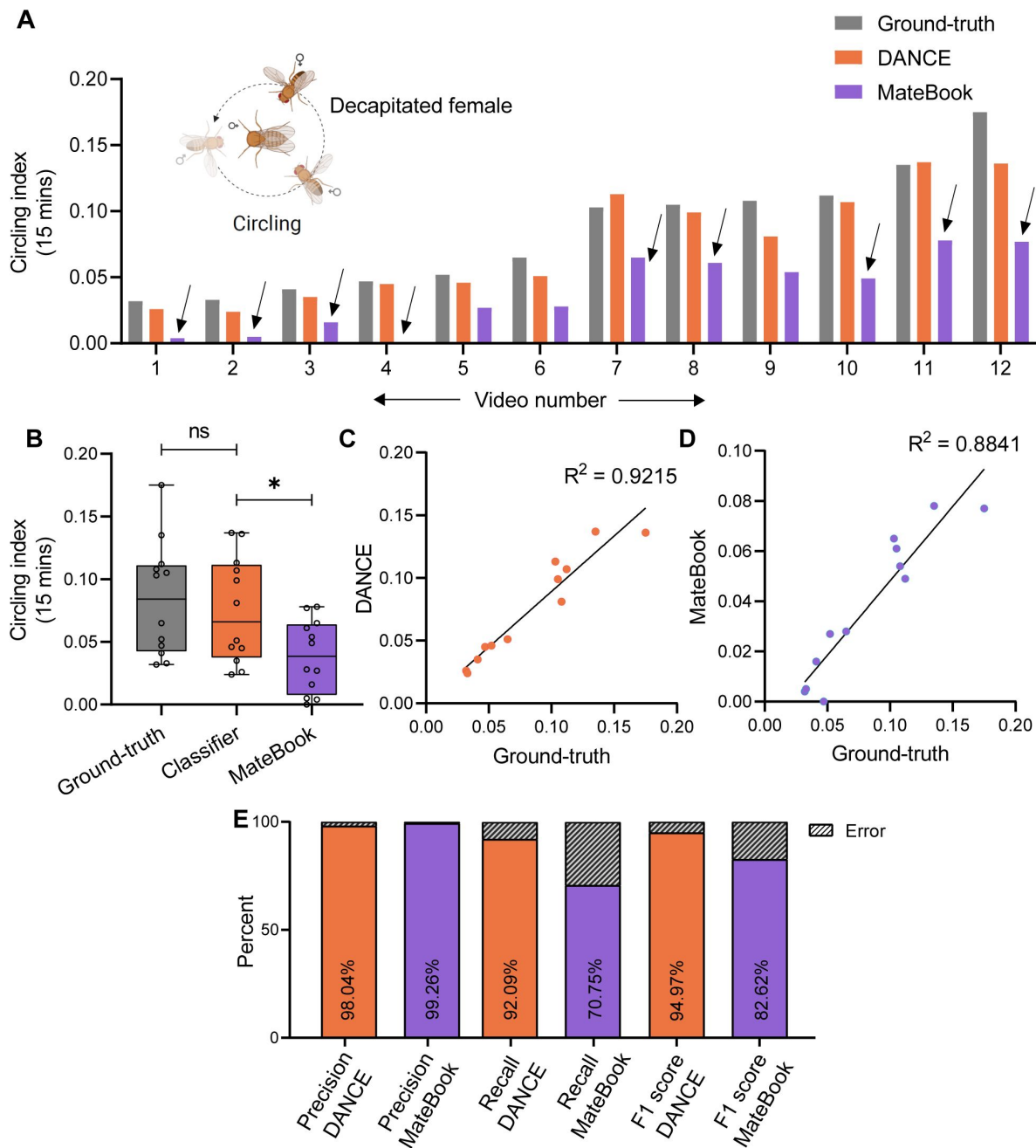


Fig. 5.

DANCE circling classifier outperforms existing methods of quantification.

(A) Circling index of males from 15 minutes long videos from the ground-truth (grey), 'DANCE circling classifier' (orange) and MateBook (purple) against decapitated virgin females. (B) Comparison of manual vs. DANCE vs. MateBook circling classifier (Ordinary one-way ANOVA with Dunnett's test, ns, $p=0.8014$; $*p=0.0157$, $n=12$). (C) Regression analysis of the circling classifier vs. ground-truth ($R^2=0.92$, $n=12$). (D) MateBook vs. ground-truth ($R^2=0.88$, $n=12$). (E) F1 score, precision, and recall of DANCE and MateBook circling classifiers against ground-truth scores.

Following

During ‘following’, the male follows around the female if it is moving in order to get closer to it and initiates the other courtship behaviors such as wing extension, licking, etc. (Spieth, 1968 [↗](#)). Since following is a relatively simpler behavior, both MateBook and the ‘DANCE following classifier’ performed well when compared to ground-truth (**Figure 5—figure supplement 2A-D** [↗](#)). However, the DANCE classifier had better precision, recall, and F1 scores (91.16%, 91.07%, and 91.12%) as compared to MateBook (65.77%, 83.39%, and 73.53%), indicating lower number of false positives and false negatives (**Figure 5—figure supplement 2E** [↗](#)).

DANCE hardware

To measure aggression and courtship in flies, existing setups (**Figure 1A** [↗](#), **Figure 1—figure supplement 1** [↗](#)-2) (Dankert et al., 2009 [↗](#); Koemans et al., 2017 [↗](#)) have several bottlenecks, making it challenging to adopt these assays by the broader neuroscience community. For instance, it requires complex and specialized hardware (Dankert et al., 2009 [↗](#); Koemans et al., 2017 [↗](#)), necessitating access to sophisticated machine shops; or 3D printers to fabricate behavioral arenas (Chowdhury et al., 2021 [↗](#); Gil-Martí et al., 2023 [↗](#)). Setting up aggression assays (Dankert et al., 2009 [↗](#); Dierick, 2007 [↗](#)) requires coating chambers with fluon to prevent flies from walking on the walls, which is tedious. Recording the videos require expensive machine vision cameras and machine vision backlights and technical expertise for data acquisition and processing. To overcome these challenges, we devised the DANCE hardware (**Figure 6** [↗](#)), which provides an inexpensive, easy, scalable, and robust alternative to existing methods of recording *Drosophila* aggression and courtship behaviors.

The DANCE hardware consists of easy to find off-the-shelf components including transparent medicine tablet foils/blister packs of different diameters, which are repurposed to serve as arenas for recording behaviors; acrylic sheets; and paper tape (**Figure 6A-D** [↗](#), Supplementary video 1-2). To substitute expensive machine vision cameras, we used widely available Android smartphone cameras for recording and expensive machine vision backlights were replaced with electronic tablets or smartphones running white light application, to provide uniform backlight/illumination source (**Figure 6E-G** [↗](#)). Aggression and courtship behaviors were recorded at 30 fps with 1080 pixel resolution. In the DANCE aggression hardware (**Figure 6A-B, E** [↗](#)) medicinal foil is slid over a base plate containing an apple juice agar food layer over which flies fight (Supplementary Video 1-2). For DANCE courtship chambers (**Figure 6C-D, F-G** [↗](#)), the medicine tablet foil is slit from the middle using a razor, through which a thin ‘separator comb’ is inserted, which is made using x-ray film. This ‘separator comb’ keeps the males and females separate until the beginning of the recording, when it is manually removed (Supplementary video 3-4).

Since the heat generated from the tablet/smartphone screen during recording can affect the aggression and courtship behaviors, we placed a transparent acrylic sheet with an air- gap of 4 mm on top of the backlight screen to facilitate heat dissipation (**Figure 6E-G** [↗](#); Supplementary Video 2, 4), which proved to be crucial to obtain robust data (**Figure 7** [↗](#)).

Benchmarking DANCE hardware

We used our DANCE classifiers to quantify aggression and courtship behaviors and found that wild type flies display these behaviors at similar levels in DANCE hardware as in the existing set ups (**Figure 7**, Supplementary video 5). We investigated the effects of social isolation vs. enrichment on aggression and courtship behaviors, since single-housing (SH) when compared to group housing (GH), increases courtship attempts (Dankert et al., 2009 [↗](#); Kim and Ehrman, 1998 [↗](#); Pan and Baker, 2014 [↗](#)), and promotes aggression (Agrawal et al., 2020 [↗](#); Wang et al., 2008 [↗](#); Yadav et

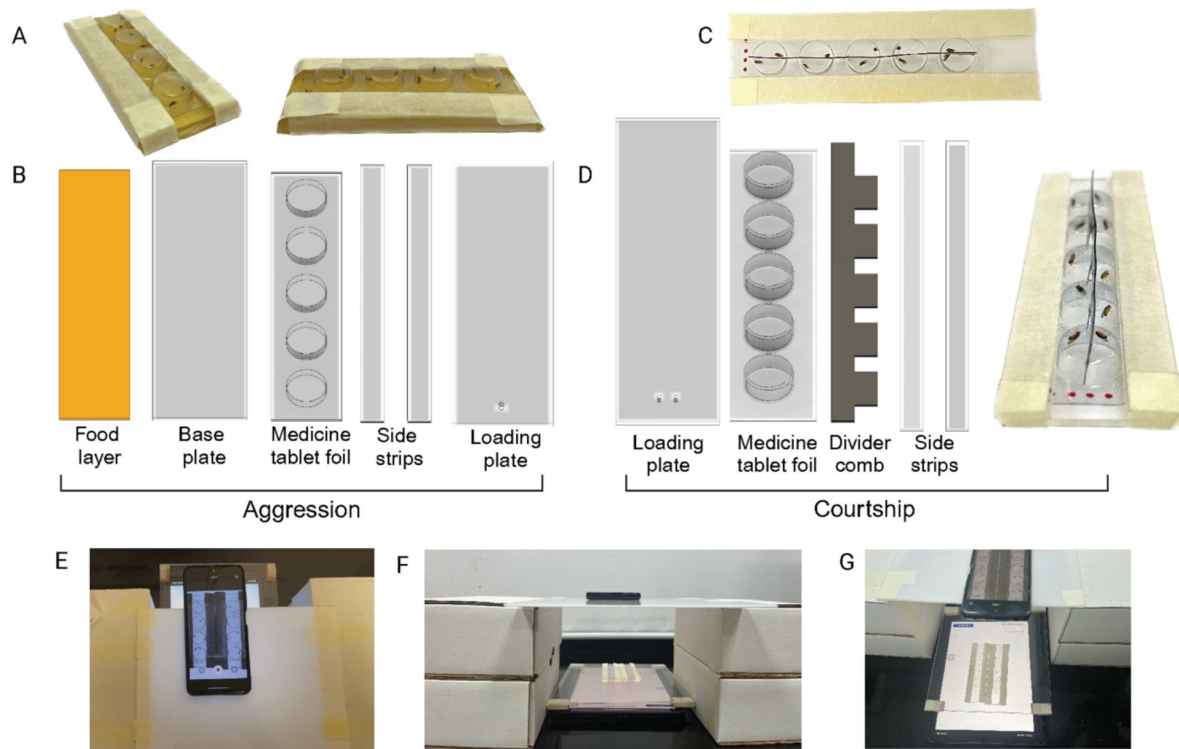


Fig. 6.

DANCE hardware and recordings setup.

(A) DANCE aggression set up (B) 3D- rendered components of the aggression set up (C) DANCE courtship set up (D) 3D- rendered components of the courtship set up, male and female are separated on either side using X-ray film separator or 'divider comb'. (E-G) Top and side views of the DANCE setup with a smartphone camera for recording and electronic tablet being used as a backlight.

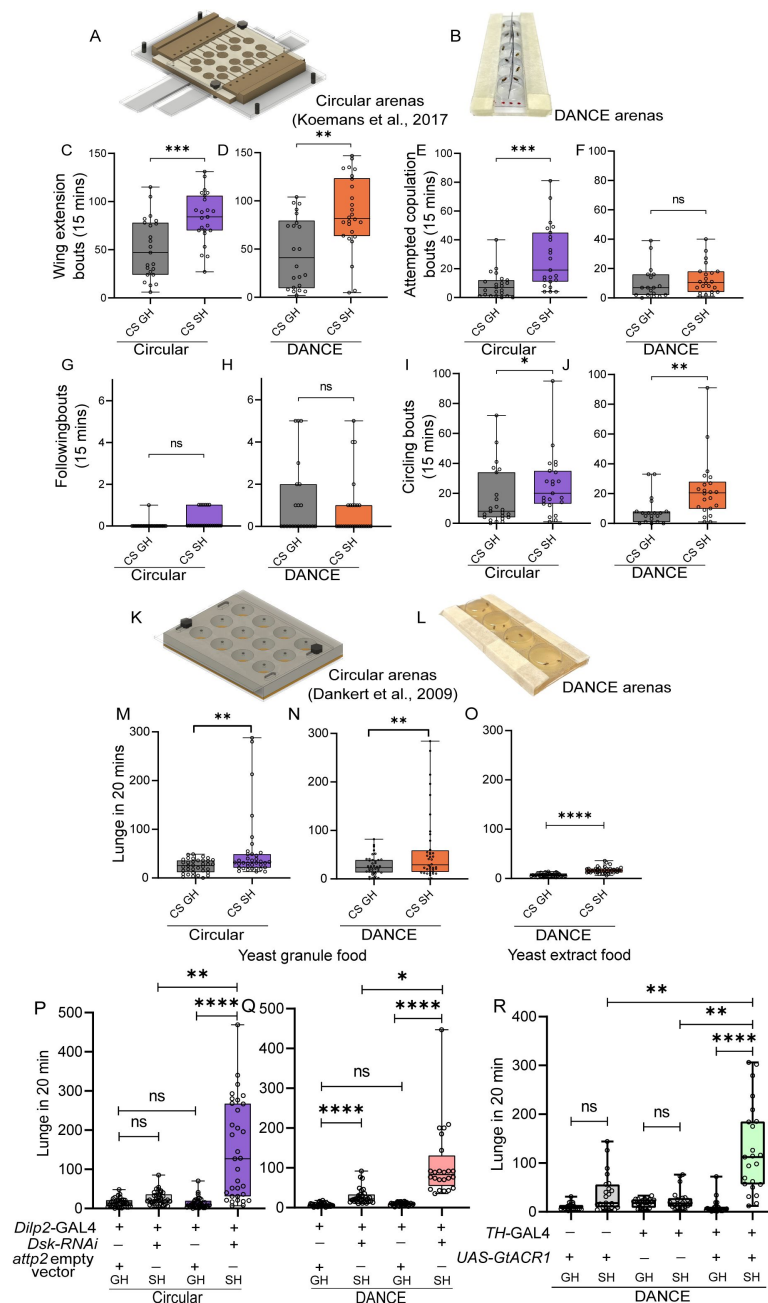


Fig. 7.

Benchmarking DANCE hardware and testing various neurogenetic tools.

(A- B) Courtship behaviors recorded using a pre-existing (circular) and DANCE set up, from group-housed (GH) and single-housed (SH) flies, for (C-D) wing extension, (C) *** $p < 0.0010$, $n = 23$; (D) ** $p < 0.0013$, GH, $n = 22$ and SH, $n = 26$. (E-F) Attempted copulation, (E) *** $p < 0.0002$, $n = 23$; (F) ns, $p < 0.1907$, GH, $n = 18$ and SH, $n = 22$. (G-H) Following, (G) ns, $p > 0.0959$, $n = 23$; (H) ns, $p < 0.6589$, GH, $n = 22$ and SH, $n = 26$. (I-J) Circling, (I) * $p < 0.012$, $n = 23$; (J) ** $p < 0.0021$, GH, $n = 19$ and SH, $n = 22$. (K-L) Aggressive lunges recorded using a pre-existing (circular) and DANCE set up. (M-N) Lunges in SH flies compared to GH flies reared on food with yeast granules, (M) ** $p < 0.0138$, $n = 36$; (N) ** $p < 0.0372$, $n = 40$. (O) Effect of yeast extract food on aggressive behavior; **** $p < 0.0001$, $n = 38$. (P-Q) Genetic knockdown of the neuropeptide Drosulfakinin (Dsk) in insulin-producing neurons using *dilp2-GAL4*. (P) ns, $p < 0.0502$, ns, $p > 0.9999$, **** $p < 0.0001$, ** $p < 0.0040$, $n = 35$. (Q) **** $p < 0.0001$, ns, $p > 0.9999$, **** $p < 0.0001$, * $p > 0.0210$, $n = 30$. (R) Optogenetic silencing of dopaminergic neurons by *UAS-GtACR1* driven by *TH-GAL4* driver, ns, $p < 0.0986$, ns, $p > 0.9999$, **** $p > 0.0001$, ** $p > 0.0012$, ** $p > 0.0013$, $n = 24$. (C-J and M-O) Mann-Whitney U test; (P-R) Kruskal-Wallis test with Dunn's multiple comparisons.

al., 2024 [↗](#)). We found comparable performance between pre-existing courtship setup (Koemans et al., 2017 [↗](#)) (**Figure 7A** [↗](#)) and the DANCE hardware (**Figure 7B** [↗](#)), both of which captured significant differences in multiple courtship behaviors in SH vs. GH males (**Figure 7C-J** [↗](#)).

Comparison of pre-existing set up for aggression (Dankert et al., 2009 [↗](#)) (**Figure 7K** [↗](#)) with the DANCE hardware (**Figure 7L** [↗](#)), showed that both the set ups faithfully captured even low levels of aggressive lunges in the wild type SH flies (**Figure 7M-N** [↗](#)). Availability of food affects *Drosophila* aggression (Lim et al., 2014 [↗](#)) and yeast in fly food was shown to interact with gut microbiome to affect male aggression (Jia et al., 2021 [↗](#)). We therefore investigated if the type of yeast used in food would affect aggression. Substituting yeast granules in fly food with yeast extract powder reduced the baseline aggression when flies were reared and differentially housed on food with yeast extract powder (**Figure 7O** [↗](#)). To investigate the effect of genetic knockdown using RNAi, which is a widely used tool in *Drosophila* genetics, we downregulated neuropeptide *Drosulfakinin* (Dsk) in insulin-producing neurons using *dilp2- GAL4*. Consistent with a previous finding (Agrawal et al., 2020 [↗](#)), aggressive lunges were significantly increased in SH males (**Figure 7P-Q** [↗](#)). Behavioral screens often utilize optogenetics to identify neuronal circuits responsible for aggression (Hoopfer et al., 2015 [↗](#); Wohl et al., 2023 [↗](#); Yadav et al., 2024 [↗](#)). It was earlier shown that while specific dopaminergic neurons may play a role in aggression, constitutive inactivation of a larger set of dopaminergic neurons using *TH-GAL4* led to unhealthy flies with defects in locomotion due to which they did not fight (Alekseyenko et al., 2013 [↗](#)). We reasoned that modern optogenetic tools may overcome the limitations of constitutive silencing and sought to investigate the role of dopaminergic neurons using our DANCE set up. We carried out optogenetic silencing during 20 minute interactions, by driving green light sensitive anion channelrhodopsin- *GtACR1* (Govorunova et al., 2015 [↗](#); Mohammad et al., 2017 [↗](#)) using *TH- GAL4*, and found significant increase in aggressive lunges in SH flies (**Figure 7R** [↗](#)). We also observed higher wing flicks as well as high intensity aggressive behaviors such as boxing and tussling (Supplementary Video 6). We did not find significant changes in locomotor activity between GH and SH flies despite 12 hours of continuous silencing (**Figure 7— figure supplement 1** [↗](#)). Together this suggests that the DANCE set up can be used with most of the common neurogenetic tools to screen for underlying molecular and neuronal circuitry governing aggression and courtship behaviors.

Discussion

Here, we present the DANCE assay, an easy to use and robust analysis pipeline and inexpensive hardware to record and quantify aggression and courtship behaviors. We developed six novel behavioral classifiers using supervised learning to accurately quantify aggression and courtship behaviors of *Drosophila* males. We developed a versatile, easy to fabricate and inexpensive hardware setup costing <\$0.30, about 10,000 fold improvement over existing set ups, using easily available, off-the-shelf components. The performance of DANCE set up is comparable with more specialized and expensive existing setups. Therefore, DANCE enables rapid behavioral screening and wider adoption by the neuroscience community including in resource limited settings such as undergraduate education.

Various components of the DANCE assay including the behavioral classifiers, hardware design and analysis codes are publicly available and can be used independently as well. This offers flexibility to the neuroscience community to further customize the DANCE classifiers for specific needs or incorporate additional data without having to develop a classifier from scratch. Our study can also serve as a template for development of future behavioral classifiers for *Drosophila* aggression such as fencing, wing flick, box, tussle, chase, female headbutt etc. (Chen et al., 2002 [↗](#); Nilsen et al., 2004 [↗](#)) and courtship behaviors, such as- tapping by male, licking the posterior abdomen of the female, female rejection etc. (Sokolowski, 2001 [↗](#)).

Such high resolution analysis of complex social interactions could give us insights into mating dynamics, sexual selection and the influence of genetics and evolution. Further, the behavioral dynamics can help uncover sequencing of dynamic components of behaviors (Nilsen et al.,

2004 [↗](#); Seeds et al., 2014 [↗](#); Simon and Heberlein, 2020 [↗](#); Zhang et al., 2020 [↗](#)). It can also serve as a template for studying complex behaviors of various species of *Drosophila* and other insects. The ease of adaptability and portability of the DANCE assay can also be useful for ethologists studying insect behaviors closer to their natural habitat.

Materials and Methods

The details of all the custom codes, analysis pipeline, sample files to run the analysis and DANCE classifiers are available in our GitHub repository at- <https://github.com/agrawallab/DANCE> [↗](#).

Fly Husbandry

Flies were reared on standard food at 25°C and 65% relative humidity with a 12hr:12hr light-dark cycle. All assays were performed at 25°C with 65% relative humidity, unless mentioned otherwise. For aggression and courtship experiments, Canton-S male flies were collected within 24 hours of eclosion and housed in a group (20 male flies per vial, 90mm length and 25mm diameter) or isolated (1 male fly per vial, 70mm length and 10mm diameter) for 6 days.

The following fly lines were acquired from the Bloomington Drosophila Stock Center (BDSC), USA: *TH-GAL4*, *Dilp2-GAL4* (RRID:BDSC_37516), *Dsk-RNAi* (RRID:BDSC_25869) and attP2 empty vector control (RRID:BDSC_36303). The *UAS-GtACR1* flies were a gift from Gaurav Das, NCCS, Pune, India and Canton-S (CS) flies were obtained from Ulrike Heberlein, HHMI, Janelia Research Campus, USA.

Aggression assay

The traditional aggression assay was performed as described previously (Dankert et al., 2009 [↗](#); Dierick, 2007 [↗](#)). In brief, the behavioral chamber is made up of 12 well-aggressive arenas (10 mm in height and 16 mm in diameter/arena). These arenas were covered by a sliding lid with 2mm loading holes to facilitate the introduction of flies. A pair of male flies that were housed either as a group (GH) or single fly (SH) were introduced into the arena wells by gentle mouth aspiration through the loading holes. After the flies were loaded, the sliding lid was tightened using screws. Fluon (Insect-a-slip, Bioquip: cat #2871B) was coated on the arena walls and was allowed to dry overnight to create a slippery surface, thereby preventing the flies from climbing the walls. Sigmacote (Sigma-Aldrich: SL2) was used to coat the sliding lid of the arena to prevent flies from walking on the arena ceiling. The chamber was placed on the food plate containing commercial apple juice (without added sugars), 2.5% w/v of sucrose, and 2.25% w/v of agarose. For experiments to test effects of fly food nutrients, either 2.4% yeast extract powder (HiMedia: RM0271) or 2.4% Yeast granules (Prime Instant Dry yeast, AB Maury, India) were mixed in fly food. For optogenetic silencing experiments, 520-540 nm wavelength of LEDS (Lumileds:2835) were used at 0.0004 μ W intensity, measured using a Power meter (Newport:843-R). For all aggression assays, flies were allowed to acclimatize in the arena for 5 minutes, and then the activity was filmed for 20 minutes. The assays were performed during ZT0 - ZT2.5 i.e. first 2.5 hours of the morning activity peak.

Courtship assay

Traditional courtship assays were performed as described by (Koemans et al., 2017 [↗](#); Ribeiro et al., 2018 [↗](#)). We introduced single pairs of male and female flies using an aspirator in an 18-well courtship chamber, consisting of individual wells/arenas with a diameter of 10 mm. Male and female flies were introduced into one half of the chamber using the sliding entry holes with a removable separator that divided the chamber into two halves. The flies were allowed to acclimatize to the arena for 5 minutes, after which the separator was removed, and the courtship behavior was recorded for 15 minutes at 30 fps using a white backlight. The assay was performed

from ZT0 to ZT3 or ZT9 to ZT12 (during peak activity windows). For generating mated females, we housed 20 females with 10 males together in a single vial for 4-6 days. For decapitated virgins, 2-4 days old virgin females were collected in a vial, and the decapitation was performed after anesthetizing on a CO₂ fly-pad, just before the assay. The assays were performed during ZT0-ZT2.5 i.e., first 2.5 hours/morning activity peak or ZT9-ZT11.5 i.e. last 2.5 hours/evening activity peak.

MateBook

To quantify courtship data, an automated software called MateBook was used (<https://github.com/Dicksonlab/MateBook>) (Ribeiro et al., 2018). MateBook tracks the fly's location and position for each video frame using the machine-vision approach. The software has built-in classifiers for each of the behaviors of a male fly during courtship, such as following, wing extension, orientation, copulation, and circling. The quantified data is generated in the form of a .tsv file, which contains bouts and other parameters related to courtship behavior. The software also provides an ethogram, which is a color-coded timeline of all the behaviors.

DANCE hardware

Circular, transparent medicinal tablet foils/blister packs were used to serve as aggression or courtship arenas, and acrylic sheets were used as base plates, and the DANCE hardware setup is held together using paper tape. The DANCE aggression arena was 13 mm in diameter and 4 mm in height. A thin acrylic base plate (2 mm thickness) with a loading hole (2 mm diameter) and a base plate with food and a side spacer were held together using paper tape (Supplementary Video 1). The tip of the food plate was also covered using paper tape for easy sliding of the medicinal tablet foil without damaging the food layer.

Sigmacote (Sigma-Aldrich: SL2) was used to coat the walls and roof of the medicine blister foil using cotton swab (Solimo, Amazon India), to prevent flies from climbing the walls of the arena. A pair of male flies were introduced into the medicinal foil wells by gentle aspiration through the loading holes. After loading flies into the medicinal foil wells, the foil was slid onto the food bed with the help of side spacers and tape on the tip of the food bed. After sliding it, the assembly was taped from ends to avoid gaps and prevent flies from escaping the arenas (Supplementary Video 2).

The DANCE courtship arena used transparent medicine blister packs with a diameter of 11mm and a height of 4.5 mm. The top surface across all arenas (5 in a row) had thin slits cut using a sharp razor to fit in the separator comb which was built by cutting used x-ray film. The foil and base plate (2 mm thickness) with loading hole (2 mm diameter) was held together by a paper tape. The base plate for courtship had two holes spaced apart in such a way that, male and female flies can be loaded on either side of the separator comb. The assembly was taped from ends to avoid gaps and prevent flies from escaping the arenas.

Once the setup was ready to record, the separator comb was gently lifted up, while ensuring the slit edges stay intact to avoid distortions during recordings (Supplementary videos 3-4).

Video acquisition

For the traditional set ups, the interaction of the flies was recorded using machine-vision cameras (DMK 33UX252 USB 3.0 monochrome camera). A white backlight (TMS, BHS4- 00-100-X-W-24V) provided the light source for both the courtship and aggression experiments. Videos were recorded at 30 frames per second (fps) for 15 minutes for courtship or 20 minutes for aggression in H.264 (.mp4) format with 1440x1080 resolution. These videos were used for training, testing, and validating the DANCE classifiers.

Alternatively, for DANCE hardware testing, we used various Android smartphone cameras (Huawei Y9 2019; OnePlus Nord CE 2 Lite 5G, model: CPH2381 or Redmi Note 11 Pro+ 5G, model: 221116SI) at 30 fps with 1080p resolution in H.264 (.mp4) format for 15 minutes (courtship) or 20 minutes (aggression) with electronic tablet (iPad Air, 5th Generation) or smartphone (iPhone 13) running a ‘white screen light app’ as the background illumination source. A transparent acrylic sheet with 4 mm spacers was kept on top of this ‘backlight’ to create an air gap to ensure heat exchange and to prevent the DANCE arenas from getting hot.

Tracking flies using FlyTracker

To successfully classify the behavior of the flies, we tracked their location and body parts using Caltech FlyTracker (Eyjolfsson et al., 2014). It allows for detection and tracking of flies based on their location, body position, orientation, and interactions across the frames. Occurrence of identity switches between male and female flies was corrected using the ‘identity correction’ tool available in the “visualizer” program of the FlyTracker package. This data was then pushed to the JAABA pipeline to develop DANCE classifiers.

Developing the DANCE behavioral classifiers

The Janelia Automatic Animal Behavior Annotator (JAABA) was used to develop DANCE behavioral classifiers. The algorithm allows experienced users to ‘encode their intuition’ to develop ‘classifiers’ that help to annotate various animal behaviors using tracking data from the behavioral recordings (Kabra et al., 2013).

Tracking data generated using FlyTracker from training videos were processed using JAABA. The stereotypic behavior bouts were labeled as ‘true behavior’, e.g., ‘Lunge’, ‘Wing extension’ etc., and an approximately equal number of obvious non-behavior bouts (‘None’), were labeled. After each round of training, the false positives predicted by the classifier would be specifically labeled as ‘None.’ This cycle of training and correcting the predictions was done until there was no improvement in the classifier’s performance. We included courtship videos of males with decapitated virgins in the training set for the attempted copulation and circling classifiers to increase the number of labeled bouts of True behavior, as these behaviors occur more frequently with decapitated virgins. Additionally, we included videos with decapitated virgins in the training sets for the following and copulation classifiers to provide more examples of the ‘None’ behavior. For the wing extension classifier, we used only mated females, as there were an adequate number of bouts representing both the true behavior and the ‘None’ behavior.

Various DANCE classifiers were developed using following frames: lunge: 2730 frames (91 seconds in total); wing extension: 99,947 frames (3,331.57 seconds in total); attempted- copulation: 39,513 frames (1,317.1 seconds in total); copulation: 56,979 frames (1899.3 seconds in total); circling: 14,396 frames (479.87 seconds in total); following: 25,787 frames (859.57 seconds in total).

DANCE classifiers, custom codes and other analysis details are available in our GitHub repository at- <https://github.com/agrawallab/DANCE>.

Characterization of male aggression and courtship behaviors

Lunge: It is defined as ‘the attacking fly rises on his hind legs, lifting his long body axis by 45° then snaps down on his opponent with his head reaching a velocity of 200 mm/s’ (Dankert et al., 2009).

Courtship classifiers were developed for five male courtship behaviors based on existing definitions (Dankert et al., 2009; Ribeiro et al., 2018). For the duration-based classifiers, a filter of minimum bout length was added in the post-processing of each classifier as a measure to

eliminate the bouts shorter than 98% of manual bouts. Details of individual courtship classifiers are mentioned below.

Wing Extension: Dankert et.al 2009, Ribeiro et al., 2018 [\[1\]](#) define a wing extension, when the angle between the body's main axis and the line from the body's center to the tip of the wing exceeds 30°, with the behavior persisting for approximately 13 frames i.e. 0.5s, at 25 fps. We processed our data using MateBook as described (Ribeiro et al., 2018 [\[1\]](#)); but with a reduced persistence filter for wing extension to 0.33 seconds from the default 0.50 seconds to match our classifier's threshold of minimum 10 frames, since in all the courtship videos that we analyzed, >98% of the bouts followed this threshold.

Attempted copulation and copulation: Existing definitions, such as (Ribeiro et al., 2018 [\[1\]](#)) define copulation as 'occlusions' persisting for more than 45 seconds and does not quantify 'attempted-copulation'. We used a similar threshold while training our classifiers but ensured that it is not merely an 'occlusion'— if the male curls its abdomen without mounting or mounts the female for a minimum of 0.33 seconds (10 frames at 30 fps) to a maximum of <45 seconds, it was considered as 'attempted-copulation' whereas mounting for >45 seconds (>1350 frames at 30 fps) was considered as 'copulation'. The results were compared with MateBook by matching its 'event settings'.

Circling: (Dankert et al., 2009 [\[1\]](#)) define circling as the fly drifting sideways in a circle with approximately constant velocity. Additionally, (Ribeiro et al., 2018 [\[1\]](#)) considered this behavior with a persistence of 13 frames (0.5s at 25 fps). We developed our 'DANCE circling classifier' based on these existing definitions, but with a reduced persistence of 10 frames (0.33s at 30 fps) and compared the results with MateBook by matching its 'event settings'.

Following: Ribeiro et al., 2018 [\[1\]](#) define the event following as when the 'the male keeps distance to the female between 2 and 5mm, while directly behind the female, while both flies are walking with a minimum speed of 2mm/s' with the behavior persisting for 25 frames i.e. 1.0s, at 25 fps. We developed our 'DANCE following extension classifier' based on these existing definitions, but with a reduced persistence of 10 frames (0.33s, at 30 fps) and compared the results with MateBook by matching its 'event settings'.

Manual behavioral annotations

To quantitatively measure the classifier's performance, manual behavioral annotations, or 'ground-truthing' were performed using the JAABA's ground-truthing mode (Kabra et al., 2013 [\[1\]](#)). The classifier's performance is measured by its robustness in predicting the behavioral bouts for the frames it was not trained on, which comprises the 'testing set.' In the testing set, we first annotate the behavioral bouts as 'true behavior' without looking at the predictions of the output. Then these videos are processed with the same classifier using JAABAPlot and compared as described subsequently.

Comparison of manual and DANCE annotations

The classifier's output was compared with that of the manual ground-truth, based on both the number of bouts for the one-frame lunge classifier and the bout duration for all duration-based courtship classifiers (wing extension, following, circling, attempted-copulation, and copulation). The duration of the bouts was used to derive the respective behavioral index. A regression analysis and classification metrics were calculated between the bout numbers or the behavior index based on the type of classifier for each of the ground-truth, DANCE, and the existing software's scores using Prism 8 (GraphPad Software).

The manual scores and JAABA output were compared, bout by bout, for the total number of true positives, false positives, and false negatives. Human annotated bouts (ground-truth) were used as a reference to the number of true positives. A ‘true positive’ is called when a JAABA bout overlaps with ground-truth bout for one or more frames (over 0.03 ms). If one ground-truth bout overlapped with two or more JAABA bouts, then all the JAABA bouts would be collectively counted as one bout, i.e., one true positive bout. Similarly, if two or more ground-truth bouts overlapped with a single JAABA bout, then the number of ground-truth bouts would be the count for the JAABA bout; i.e., two or more number of true positives.

When a JAABA bout does not overlap with any ground-truth (human annotation), then it would be considered a false positive. When a ground-truth bout does not overlap with any JAABA bout, it is considered a false negative (Leng et al., 2020 [DOI](#)).

We used a custom Python code, to calculate the number of total overlaps between the ground-truth, the DANCE classifier’s, and the existing algorithm’s output. Once we had the total number of true positives, false positives, and false negatives, we derived the precision, recall and F1 scores to gain a deeper understanding of the performance of DANCE classifiers. Precision rate represents the ratio of correctly predicted positive observations to the total predicted positives; the recall rate represents the ratio of correctly predicted positive observations to all actual positives; and the F1 score is a single metric which is a harmonic mean of both precision and recall. We used the following formulas to calculate precision, recall, and F1 score:

$$\begin{aligned} \text{precision} &= \frac{(\text{number of true positives})}{(\text{number of true positives}) + (\text{number of false positives})} \\ \text{recall} &= \frac{(\text{number of true positives})}{(\text{number of true positives}) + (\text{number of false negatives})} \\ \text{F1 score} &= 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \end{aligned}$$

Statistical analysis

For non-normally distributed data, Mann-Whitney U test or ANOVA with appropriate posthoc correction was used. Statistical analysis of behavioral data was performed using Prism 8 (GraphPad Software), custom scripts or Microsoft Excel.

Acknowledgements

We thank Barry Dickson from Queensland Brain Institute, Australia for encouraging discussion and suggestions related to MateBook and Ben Arthur, HHMI, Janelia Research Campus, USA for helpful suggestions for setting up the MateBook analysis. We thank Mayank Kabra, Kristin Branson and colleagues for developing JAABA and providing excellent support to the user community. We thank Ulrike Heberlein, HHMI, Janelia Research Campus, USA and Gaurav Das, NCCS, Pune and Bloomington Drosophila Stock Center (NIH P40OD018537) for fly stocks. We thank Santhosh Chidangil, MAHE for help with LED power meter. We thank Santosh D’Mello, LSU Shreveport and Gaurav Das for helpful discussions; Gaurav Das and Toshiharu Ichinose for critical reading and feedback on the manuscript.

Additional information

Figure 1—Figure Supplement 1.

Aggression chamber described by Dankert *et al.*, 2009

It consists of a bottom food plate, 12 wells (aggression arenas), a top plate with fly loading holes and a screw slot for sliding the loading plate.

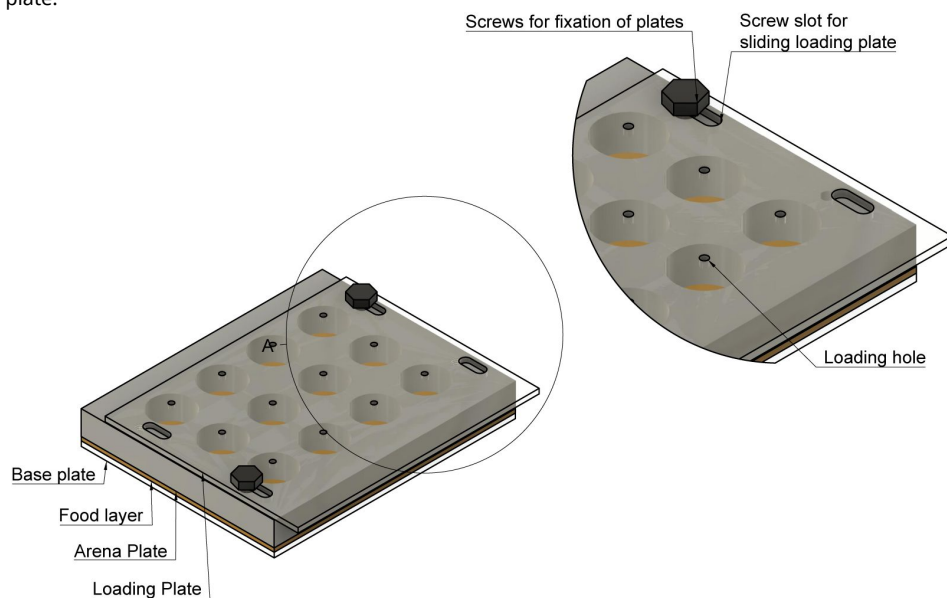
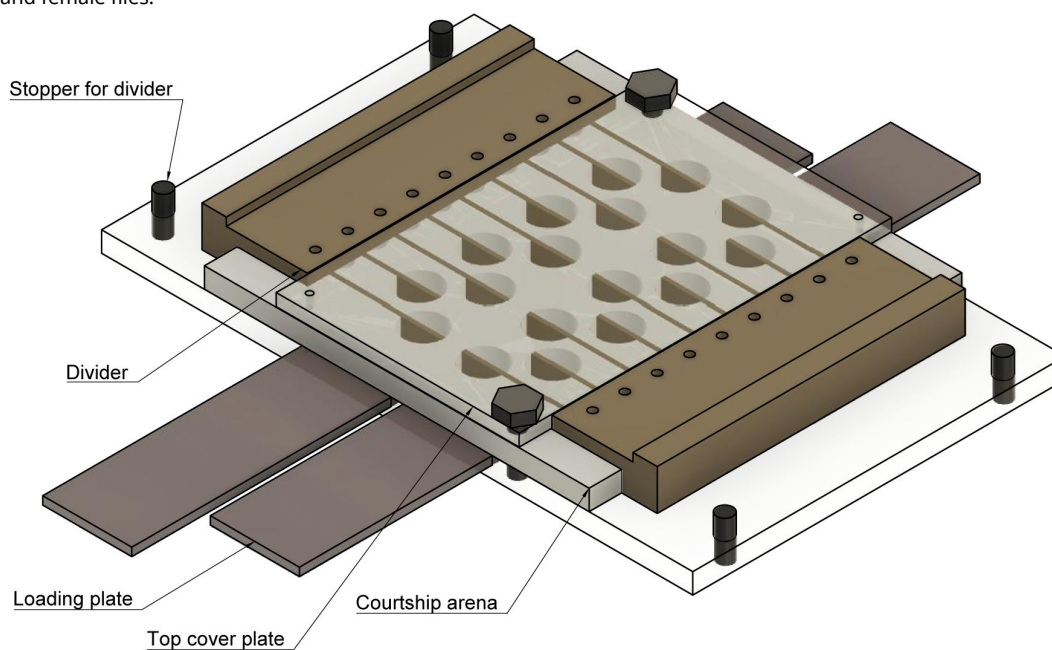


Figure 1—Figure Supplement 2.

Courtship set-up described by Koemans *et al.*, 2017

It consists of 18 wells (courtship arenas), a top cover plate, a sliding loading plate, and a sliding divider assembly to separate male and female flies.



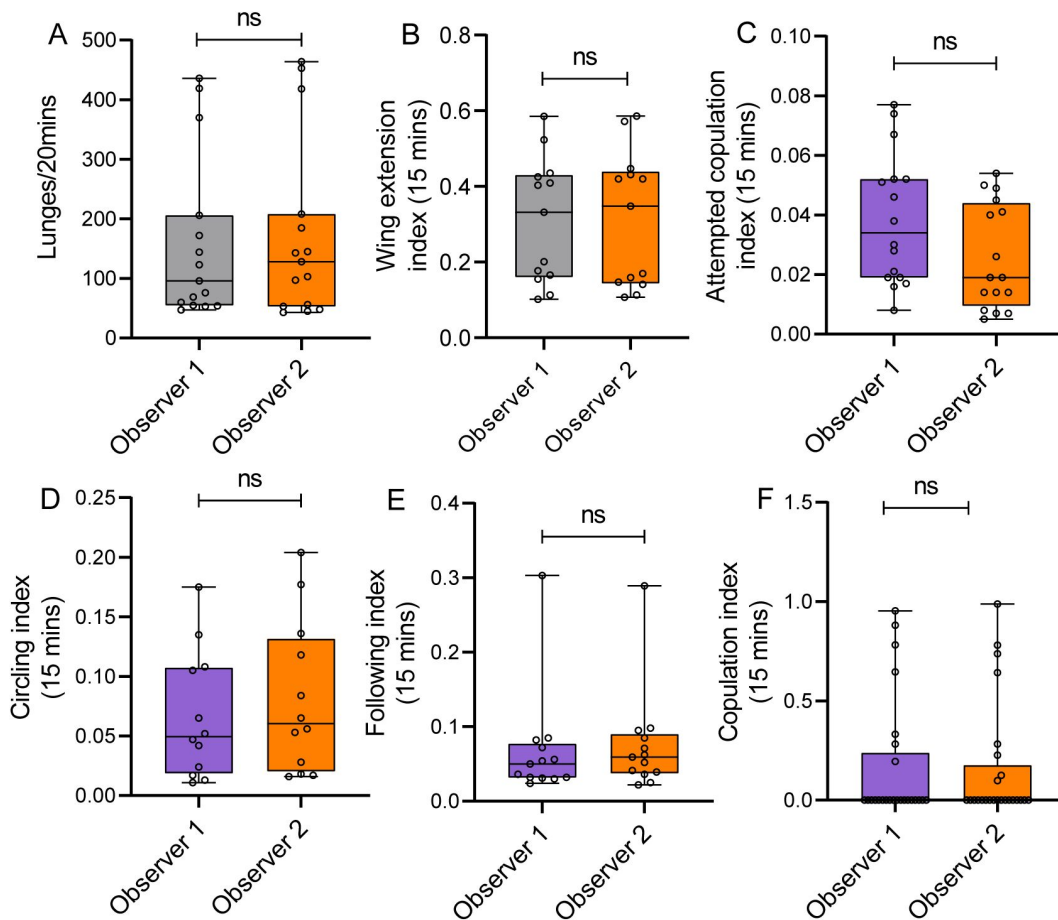


Figure 2—Figure Supplement 1.

Comparison of annotations by two independent evaluators to assess observer bias during classifier ground-truthing.

(A) Aggressive lunge, $p=0.8789$, $n=15$. (B) Courtship including wing extension, $p=0.9999$, $n=13$. (C) Attempted copulation, $p=0.0571$, $n=16$. (D) Circling, $p=0.4343$, $n=12$. (E) Following, $p=0.4405$, $n=13$. (F) Copulation, $p=0.9221$, $n=25$. (A-F) Mann-Whitney U test.

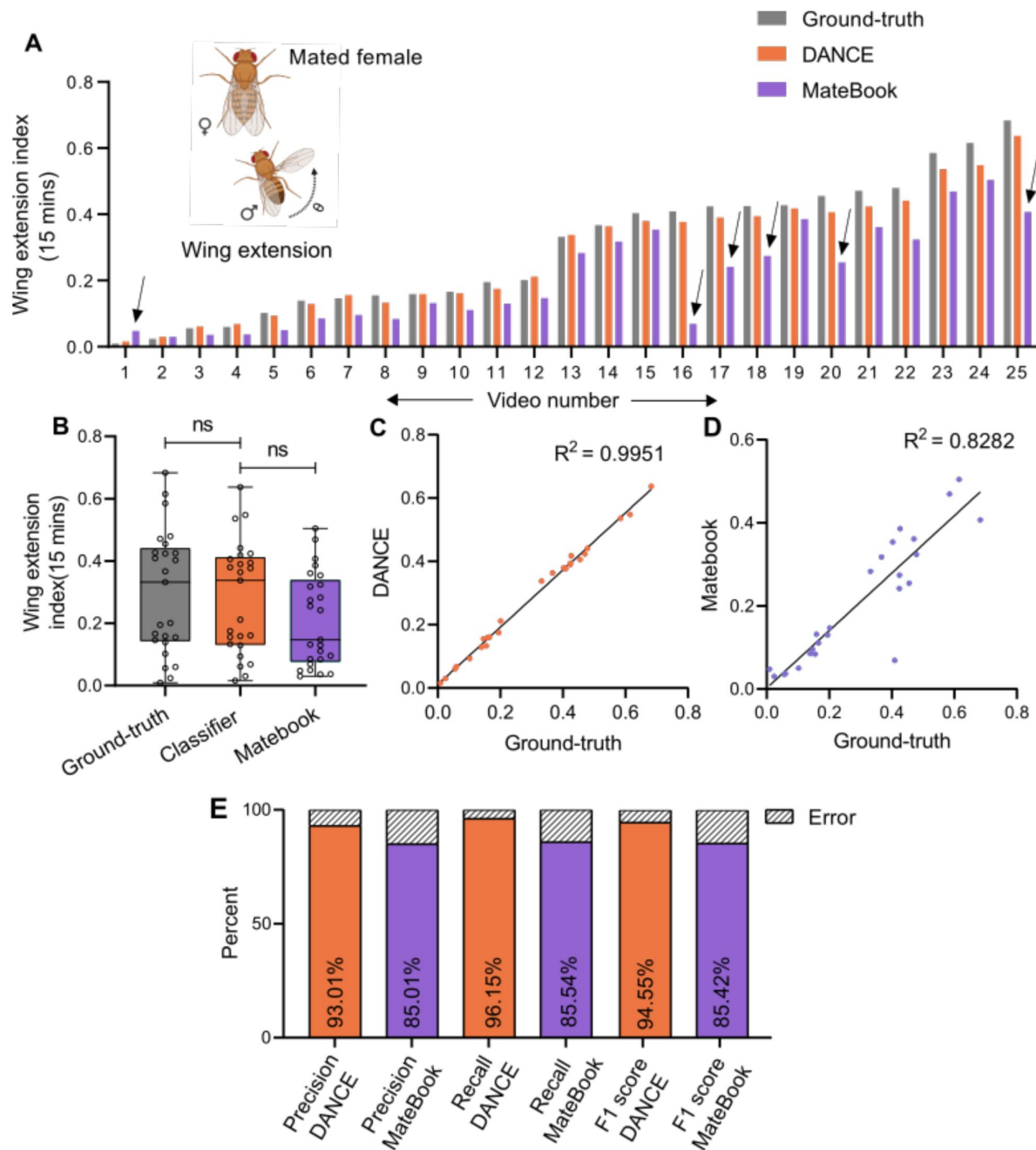


Figure 3—Figure Supplement 1.

DANCE wing extension classifier outperforms existing quantification methods in videos with mated females.

(A) Wing extension of males from 15-minute-long videos from the ground-truth (grey), the DANCE classifier (orange) and MateBook (purple) against mated females. MateBook underscored wing extension across multiple videos (black arrows) (Friedman's ANOVA with Dunn's test, $p=0.3582$; $p<0.0001$, $n=25$). (B) Comparison of ground-truth vs. DANCE vs. MateBook wing extension classifier (Kruskal-Wallis ANOVA with Dunn's test, $p>0.9999$; $p=0.1039$, $n=25$). (C) Regression analysis of the DANCE classifier vs. ground-truth ($R^2=0.9951$, $n=25$). (D) MateBook vs. ground-truth ($R^2=0.8282$, $n=25$). (E) F1 score, precision, and recall of DANCE classifier and MateBook against ground-truth scores.

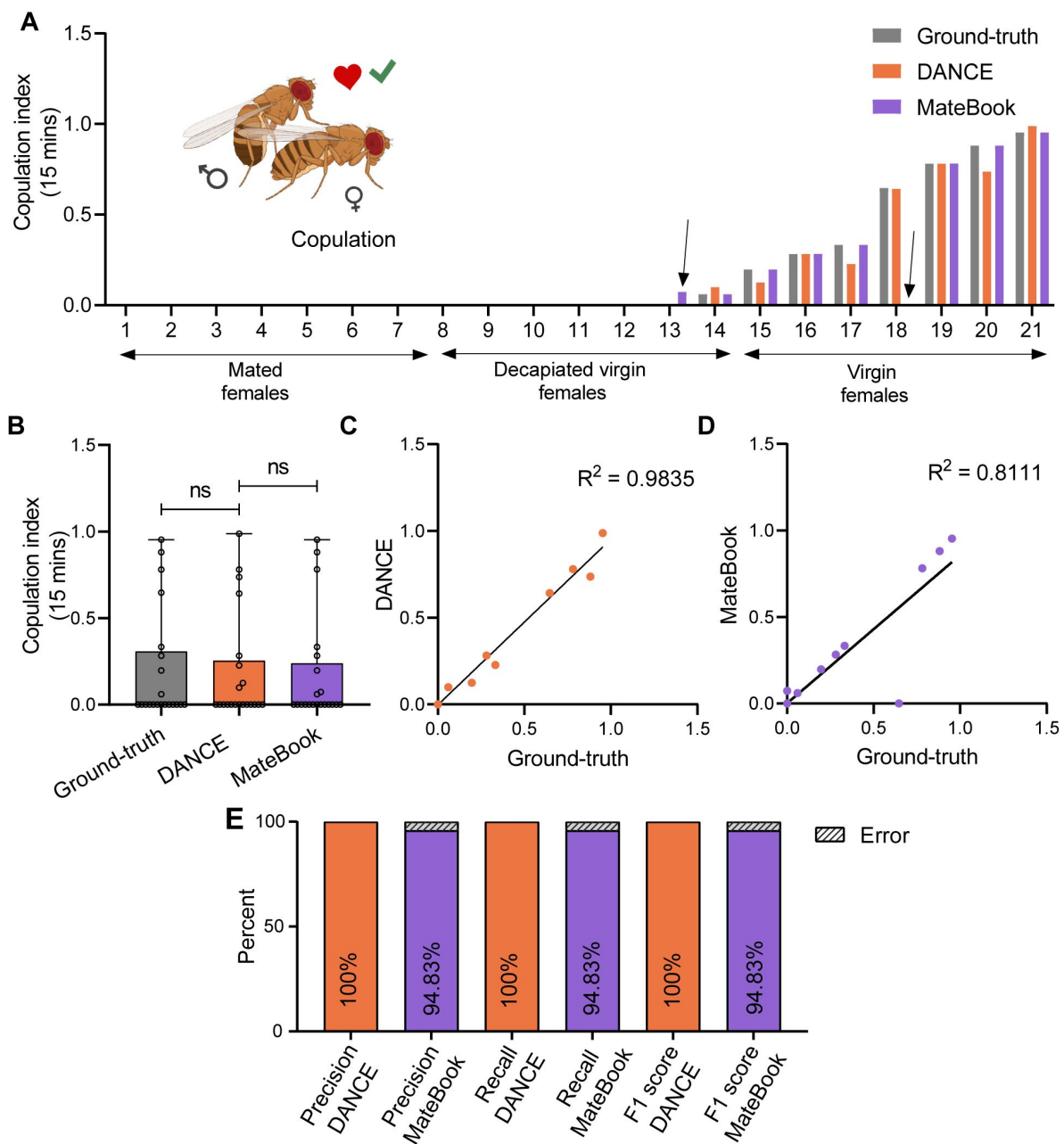


Figure 4—Figure Supplement 1.

DANCE copulation classifier evaluation in mixed female dataset.

(A) Quantification of copulation in 15 minutes from individual videos showing scores from the manual method (grey), DANCE copulation classifier (orange), and MateBook (purple) (Friedman ANOVA with Dunn's test, $p > 0.9999$; $p > 0.9999$, $n = 21$). (B) Box plot comparison of manual vs. DANCE copulation classifier vs. MateBook (Kruskal-Wallis ANOVA with Dunn's test, $p > 0.9999$; $p > 0.9999$, $n = 21$). (C) Bar plots showing F1 score, precision, and recall of DANCE copulation classifier and MateBook against ground-truth scores. (D) Regression analysis of the copulation classifier vs. manual scores ($R^2 = 0.98$, $n = 21$). (E) MateBook vs. manual scores ($R^2 = 0.81$, $n = 21$).

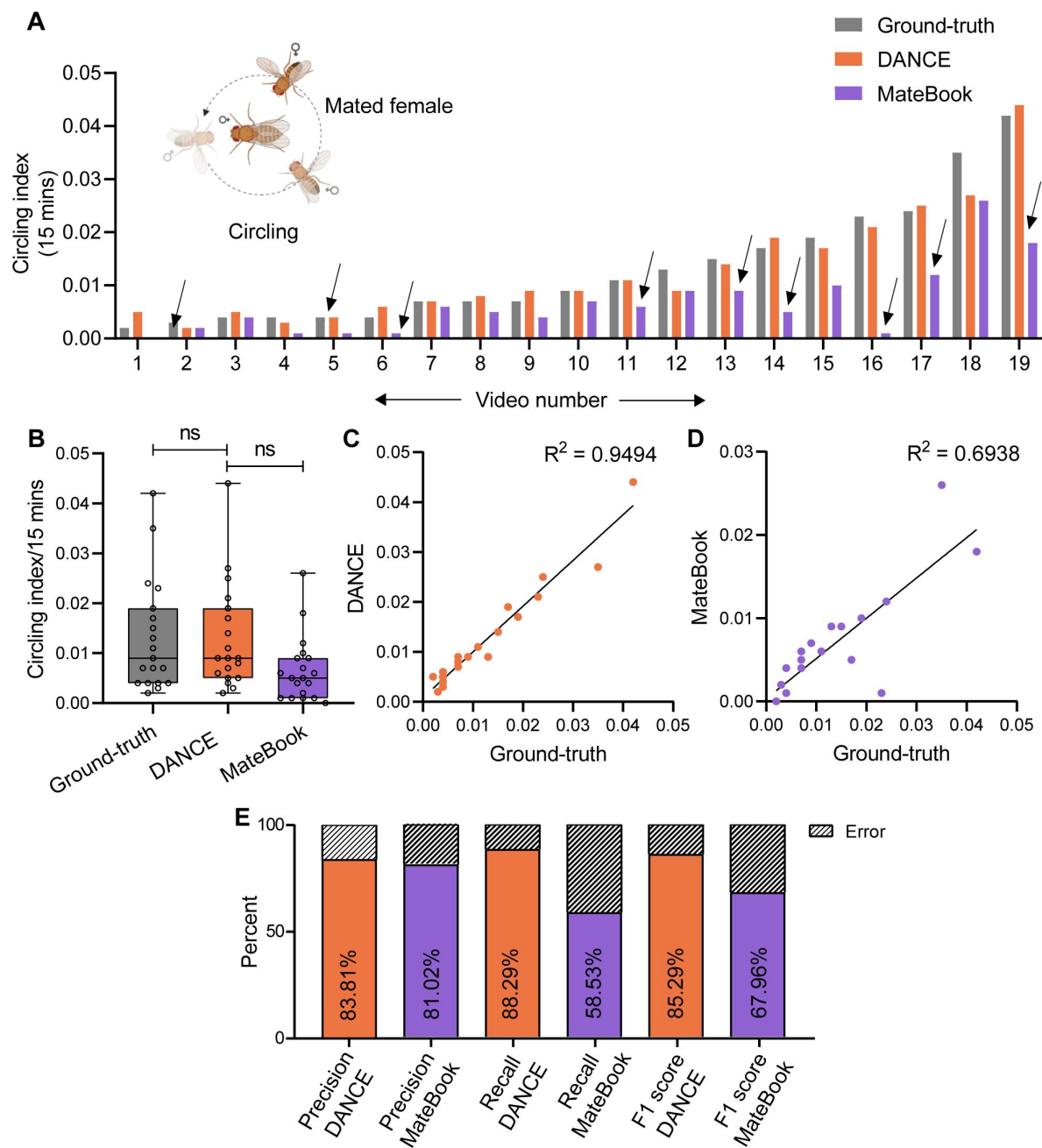


Figure 5—Figure Supplement 1.

DANCE circling classifier evaluation in the mated female dataset.

(A) Circling index of males from 15-minute videos from the ground-truth (grey), 'DANCE circling classifier' (orange) and MateBook (purple) against mated female dataset (Friedman's ANOVA with Dunn's test, $p > 0.9999$; $p < 0.0001$, $n = 19$). (B) Comparison of manual vs. DANCE vs. MateBook circling classifier (Kruskal-Wallis ANOVA, $p > 0.9999$; $p = 0.0822$, $n = 19$). (C) Regression analysis of the circling DANCE classifier vs. ground-truth ($R^2 = 0.9494$, $n = 19$). (D) MateBook vs. ground-truth ($R^2 = 0.6938$, $n = 19$). (E) F1 score, precision, and recall of DANCE and MateBook circling classifiers against ground-truth scores.

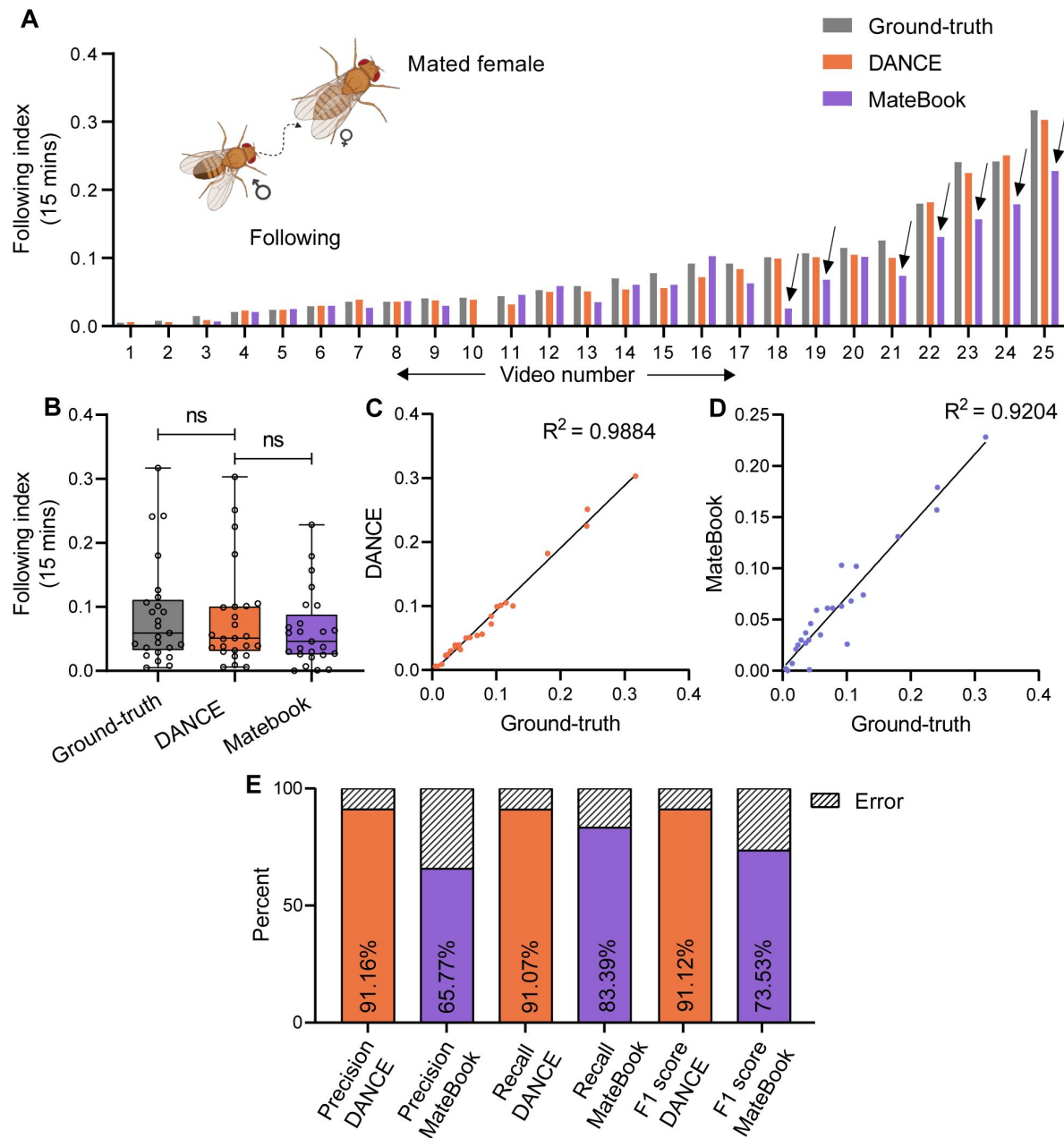


Figure 5—Figure Supplement 2.

DANCE following classifier evaluation in the mated female dataset.

(A) Following index of males from 15 minutes long videos from the ground- truth (grey), 'DANCE following classifier' (orange) and MateBook (purple) against mated females (Friedman's ANOVA with Dunn's test, $p=0.1794$; $p=0.0029$, $n=25$). (B) Box plot comparison of manual vs. DANCE following classifier vs. MateBook (Kruskal-Wallis ANOVA with Dunn's test, $p>0.9999$; $p=0.5287$, $n=25$). (C) Regression analysis of the following classifier vs. ground-truth ($R^2=0.9894$, $n=25$). (D) MateBook vs. ground-truth ($R^2=0.9204$, $n=25$). (E) F1 score, precision, and recall of DANCE and MateBook following classifiers against ground-truth scores.

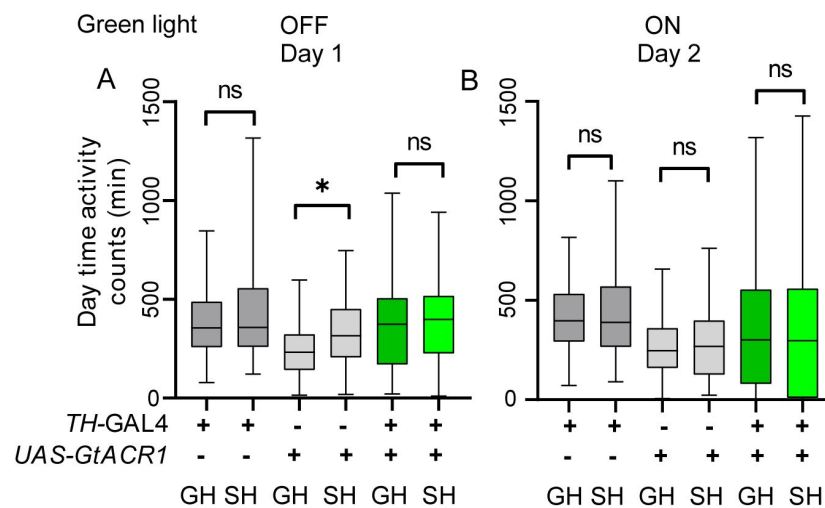


Figure 7—Figure Supplement 1.

Effect of optogenetic silencing of dopaminergic neurons on *Drosophila* activity.

(A, B) Transient silencing of dopaminergic neurons using *UAS-GtACR1* driver didn't affect daytime activity between SH and GH flies. (A) Without silencing mediated by green light on day 1; *TH-GAL4*, ns, $p=0.7383$, GH: $n=43$, SH: $n=43$; *UAS-GtACR1*, ns, $p=0.4812$, GH: $n=51$, SH: $n=42$; *TH-GAL4>UAS-GtACR1*, ns, $p=0.9942$, GH: $n=54$, SH: $n=51$. (B) With silencing mediated by green light on day 2; *TH-GAL4*, ns, $p=0.9976$, GH: $n=43$, SH: $n=43$; *UAS-GtACR1*, ns, $p=0.9779$, GH: $n=51$, SH: $n=42$; *TH-GAL4>UAS-GtACR1*, ns, $p=0.9974$, GH: $n=54$, SH: $n=51$. One-way ANOVA with Tukey's multiple comparisons test for within each day comparison. Two-way ANOVA for comparison across days; interaction, ns, $p=0.5504$, silencing, ns, $p=0.5172$, housing, ns, $p=0.1602$, GH: $n=43$, SH: $n=41$; *UAS-GtACR1*, interaction, ns, $p=0.4533$, silencing, ns, $p=0.3602$, housing, $*p=0.255$, GH: $n=51$, SH: $n=42$; *TH-GAL4>UAS-GtACR1*, interaction, ns, $p=0.9977$, silencing, ns, $p=0.2454$, housing, ns, $p=0.5868$, GH: $n=54$, SH: $n=51$.

Funding

This work was supported by funding to PA from the Department of Biotechnology (DBT), Ramalingaswami Re-entry Fellowship (BT/RLF, Re-entry/34/2018), and DBT, Research grant (BT/PR36166/BRB/10/1859/2020) by the DBT, Ministry of Science and Technology, Government of India. RSPY was supported by a TMA Pai fellowship from MAHE and DBT, Research grant to PA. FA was supported by Ramalingaswami fellowship, DBT, India to PA. TK was supported by DBT, Research grant to PA. MV is supported by NFST, Ministry of Tribal Affairs, Government of India. SA is supported by the Department of Science and Technology (DST) Inspire Fellowship. PPP is supported by a TMA Pai fellowship from MAHE, India. SBS is supported by the Anusandhan National Research Foundation (ANRF), Ministry of Science and Technology, Government of India, grant to PA (CRG/2022/006846).

Additional files

Supplementary File 1. Bill of material for DANCE set up and comparison with existing setups. [↗](#)

Supplementary File 2. Comparison between DANCE lunge classifier, ground-truth and existing methods. [↗](#)

Supplementary File 3. Definitions for behavioral classifiers. [↗](#)

Supplementary File 4. Comparison between DANCE courtship classifiers, ground-truth and MateBook. [↗](#)

Supplementary Video 1. 3D rendered DANCE Aggression hardware. [↗](#)

Supplementary Video 2. Using DANCE hardware setup for recording aggression. [↗](#)

Supplementary Video 3. 3D rendered DANCE Courtship hardware. [↗](#)

Supplementary Video 4. Using DANCE hardware setup for recording Courtship. [↗](#)

Supplementary Video 5. Aggression and courtship behaviors recorded in DANCE hardware. [↗](#)

Supplementary Video 6. Optogenetic silencing of dopaminergic neurons in DANCE shows increased aggression. [↗](#)

References

1. Agrawal P, Kao D, Chung P, Looger LL (2020) **The neuropeptide Drosulfakinin regulates social isolation-induced aggression in *Drosophila*** *J Exp Biol* **223**:ejeb207407
2. Alekseyenko OV, Chan Y-B, Li R, Kravitz EA (2013) **Single dopaminergic neurons that modulate aggression in *Drosophila*** *Proc Natl Acad Sci U S A* **110**:6151–6156
3. Anderson DJ, Perona P (2014) **Toward a Science of Computational Ethology** *Neuron* **84**:18–31
4. Asahina K (2017) **Neuromodulation and Strategic Action Choice in *Drosophila* Aggression** *Annu Rev Neurosci* **40**:51–75
5. Asahina K, Watanabe K, Duistermars BJ, Hoopfer E, González CR, Eyjólfsson EA, Perona P, Anderson DJ (2014) **Tachykinin-Expressing Neurons Control Male-Specific Aggressive Arousal in *Drosophila*** *Cell* **156**:221–235
6. Bastock M, Manning A (1955) **The Courtship of *Drosophila Melanogaster*** *Behaviour* **8**:85–110
7. Bennet-Clark HC, Ewing AW (1969) **Pulse interval as a critical parameter in the courtship song of *Drosophila melanogaster*** *Anim Behav* **17**:755–759
8. Benzer S (1967) **Behavioral mutants of *Drosophila* isolated by countercurrent distribution** *Proc Natl Acad Sci U S A* **58**:1112–1119
9. Branson K, Robie AA, Bender J, Perona P, Dickinson MH (2009) **High-throughput ethomics in large groups of *Drosophila*** *Nat Methods* **6**:451–457
10. Brown RGB (1965) **Courtship behaviour in the *Drosophila* obscura group. II. Comparative studies** *Behaviour* **25**:281–322
11. Chen S, Lee AY, Bowens NM, Huber R, Kravitz EA (2002) **Fighting fruit flies: a model system for the study of aggression** *Proc Natl Acad Sci U S A* **99**:5664–5668
12. Chiu H, Hoopfer ED, Coughlan ML, Pavlou HJ, Goodwin SF, Anderson DJ (2021) **A circuit logic for sexually shared and dimorphic aggressive behaviors in *Drosophila*** *Cell* **184**:507–520
13. Chowdhury B, Wang M, Gnerer JP, Dierick HA (2021) **The Divider Assay is a high-throughput pipeline for aggression analysis in *Drosophila*** *Communications Biology* **4**:1–12
14. Cook R, Cook A (1975) **The attractiveness to males of female *Drosophila melanogaster*: Effects of mating, age and diet** *Anim Behav* **23**:521–526
15. Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P (2009) **Automated monitoring and analysis of social behavior in *Drosophila*** *Nat Methods* **6**:297–303
16. Datta SR, Anderson DJ, Branson K, Perona P, Leifer A (2019) **Computational Neuroethology: A Call to Action** *Neuron* **104**:11–24

17. Davis SM, Thomas AL, Liu L, Campbell IM, Dierick HA (2018) **Isolation of aggressive behavior mutants in drosophila using a screen for wing damage** *Genetics* **208**:273–282
18. Davis SM, Thomas AL, Nomie KJ, Huang L, Dierick HA (2014) **Tailless and Atrophin control Drosophila aggression by regulating neuropeptide signalling in the pars intercerebralis** *Nat Commun* **5**:3177
19. Dickson BJ (2008) **Wired for sex: The neurobiology of Drosophila mating decisions** *Science* **322**:904–909
20. Dierick HA (2007) **A method for quantifying aggression in male drosophila melanogaster** *Nat Protoc* **2**:2712–2718
21. Dow MA, von Schilcher F (1975) **Aggression and mating success in Drosophila melanogaster** *Nature* **254**:511–512
22. Eyjolfsson E, Branson S, Burgos-Artizzu XP, Hoopfer ED, Schor J, Anderson DJ, Perona P (2014) **Detecting Social Actions of Fruit Flies** *Computer Vision – ECCV 2014* :772–787
23. Fowler GL (1973) **Some Aspects of the Reproductive Biology of Drosophila: Sperm Transfer, Sperm Storage, and Sperm Utilization** In: Caspari EW, editors. *Advances in Genetics* Academic Press pp. 293–360
24. Gill KS (1963) **A mutation causing abnormal courtship and mating behavior in males of Drosophila melanogaster** *Am Zool* **3**:507
25. Gil-Martí B, Barredo CG, Pina-Flores S, Poza-Rodríguez A, Treves G, Rodríguez-Navas C, Camacho L, Pérez-Serna A, Jiménez I, Brazales L, Fernández J, Martín FA (2023) **A simplified courtship conditioning protocol to test learning and memory in Drosophila** *STAR Protoc* **4**:101572
26. Gomez-Marin A, Paton JJ, Kampff AR, Costa RM, Mainen ZF (2014) **Big behavioral data: psychology, ethology and the foundations of neuroscience** *Nat Neurosci* **17**:1455–1462
27. Govorunova EG, Sineshchikov OA, Janz R, Liu X, Spudich JL (2015) **Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics** *Science* **349**:647–650
28. Greenspan RJ, Ferveur J-F (2000) **Courtship in Drosophila** *Annu Rev Genet* **34**:205–232
29. Hall JC (2002) **Courtship lite: a personal history of reproductive behavioral neurogenetics in Drosophila** *J Neurogenet* **16**:135–163
30. Hall JC (1978) **Courtship among males due to a male-sterile mutation in Drosophila melanogaster** *Behav Genet* **8**:125–141
31. Hoffmann AA (1987a) **A laboratory study of male territoriality in the sibling species Drosophila melanogaster and D. simulans** *Anim Behav* **35**:807–818
32. Hoffmann AA (1987b) **Territorial Encounters Between Drosophila Males of Different Sizes** *Anim Behav* **1**:70–71
33. Hoopfer ED, Jung Y, Inagaki HK, Rubin GM, Anderson DJ (2015) **P1 interneurons promote a persistent internal state that enhances inter-male aggression in Drosophila** *eLife* **4**:1–27

34. Hoyer SC, Eckart A, Herrel A, Zars T, Fischer SA, Hardie SL, Heisenberg M. (2008) **Octopamine in male aggression of *Drosophila*** *Curr Biol* **18**:159–167
35. Ishii K, Cortese M, Leng X, Shokhirev MN, Asahina K (2022) **A neurogenetic mechanism of experience-dependent suppression of aggression** *Sci Adv* **8**:eabg3203
36. Jacobs ME (1960) **Influence of Light on Mating of *Drosophila Melanogaster*** *Ecology* **41**:182–188
37. Jia Y, Jin S, Hu K, Geng L, Han C, Kang R, Pang Y, Ling E, Tan EK, Pan Y, Liu W (2021) **Gut microbiome modulates *Drosophila* aggression through octopamine signaling** *Nat Commun* **12** <https://doi.org/10.1038/s41467-021-23041-y>
38. Jung Y, Kennedy A, Chiu H, Mohammad F, Claridge-Chang A, Anderson DJ (2020) **Neurons that Function within an Integrator to Promote a Persistent Behavioral State in *Drosophila*** *Neuron* **105**:322–333
39. 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
40. Kessler S (1962) **Courtship rituals and reproductive isolation between the races or incipient species of *Drosophila paulistorum*** *Am Nat* **96**:117–121
41. Kim Y-K, Ehrman L (1998) **Developmental Isolation and Subsequent Adult Behavior of *Drosophila paulistorum*. IV. Courtship** *Behav Genet* **28**:57–65
42. Kim Y-K, Saver M, Simon J, Kent CF, Shao L, Eddison M, Agrawal P, Texada M, Truman JW, Heberlein U (2018) **Repetitive aggressive encounters generate a long-lasting internal state in *Drosophila melanogaster* males** *Proceedings of the National Academy of Sciences* **115**:1099–1104
43. Koemans TS, Oppitz C, Donders RAT, van Bokhoven H, Schenck A, Keleman K, Kramer JM (2017) **Courtship Conditioning As a Measure of Learning and Memory** *J Vis Exp* <https://doi.org/10.3791/55808>
44. Kravitz EA, Fernandez MP (2015) **Aggression in *Drosophila*** *Behav Neurosci* **129**:549–563
45. Leng X, Wohl M, Ishii K, Nayak P, Asahina K (2020) **Quantifying influence of human choice on the automated detection of *Drosophila* behavior by a supervised machine learning algorithm** *PLoS One* **15**:e0241696
46. Lim RS, Eyjólfsson E, Shin E, Perona P, Anderson DJ (2014) **How food controls aggression in *Drosophila*** *PLoS One* **9**
47. MacBean IT, Parsons PA (1967) **Directional selection for duration of copulation in *Drosophila melanogaster*** *Genetics* **56**:233–239
48. Markow TA (1987) **Behavioral and sensory basis of courtship success in *Drosophila melanogaster*** *Proceedings of the National Academy of Sciences* **84**:6200–6204
49. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M (2018) **DeepLabCut: markerless pose estimation of user-defined body parts with deep learning** *Nat Neurosci* **21**:1281–1289

50. Mohammad F, Stewart JC, Ott S, Chlebikova K, Chua JY, Koh TW, Ho J, Claridge-Chang A (2017) **Optogenetic inhibition of behavior with anion channelrhodopsins** *Nat Methods* **14**:271–274
51. Nilsen SP, Chan Y-B, Huber R, Kravitz EA (2004) **Gender-selective patterns of aggressive behavior in *Drosophila melanogaster*** *Proc Natl Acad Sci U S A* **101**:12342–12347
52. O'Dell KMC (2003) **The voyeurs' guide to *Drosophila melanogaster* courtship** *Behav Processes* **64**:211–223
53. Pan Y, Baker BS (2014) **Genetic Identification and Separation of Innate and Experience-Dependent Courtship Behaviors in *Drosophila*** *Cell* **156**:236–248
54. Pantalia M, Lin Z, Tener SJ, Qiao B, Tang G, Ulgherait M, O'Connor R, Delventhal R, Volpi J, Syed S, Itzhak N, Canman JC, Fernández MP, Shirasu-Hiza M (2023) ***Drosophila* mutants lacking the glial neurotransmitter-modifying enzyme Ebony exhibit low neurotransmitter levels and altered behavior** *Sci Rep* **13**:10411
55. Pavlou HJ, Goodwin SF (2013) **Courtship behavior in *Drosophila melanogaster*: towards a 'courtship connectome'** *Curr Opin Neurobiol* **23**:76–83
56. Reza MA, Mhatre SD, Morrison JC, Utreja S, Saunders AJ, Breen DE, Marendra DR (2013) **Automated analysis of courtship suppression learning and memory in *Drosophila melanogaster*** *Fly* **7**:105–111
57. Ribeiro IMA, Drews M, Bahl A, Machacek C, Borst A, Dickson BJ (2018) **Visual Projection Neurons Mediating Directed Courtship in *Drosophila*** *Cell* **174**:607–621
58. Robie AA, Hirokawa J, Edwards AW, Simpson JH, Reiser MB, Branson K, Robie AA, Hirokawa J, Edwards AW, Umayam LA, Lee A, Phillips ML, Card GM, Korff W, Rubin GM, Simpson JH, Reiser MB, Branson K (2017a) **Mapping the Neural Substrates of Behavior Resource Mapping the Neural Substrates of Behavior** *Cell* **170**:393–406
59. Robie AA, Seagraves KM, Egnor SER, Branson K (2017b) **Machine vision methods for analyzing social interactions** *J Exp Biol* **220**:25–34
60. Seeds AM, Ravbar P, Chung P, Hampel S, Jr Midgley FM, Mensh BD, Simpson JH (2014) **A suppression hierarchy among competing motor programs drives sequential grooming in *Drosophila*** *eLife* **3**:e02951
61. Shorey HH (1962) **Nature of the Sound Produced by *Drosophila melanogaster* during Courtship** *Science* **137**:677–678
62. Simon JC, Heberlein U (2020) **Social hierarchy is established and maintained with distinct acts of aggression in male *Drosophila melanogaster*** *J Exp Biol* **223**:ejeb232439
63. Sokolowski MB (2001) ***Drosophila*: genetics meets behaviour** *Nat Rev Genet* **2**:879–890
64. Spieth HT (1974) **Courtship behavior in *Drosophila*** *Annu Rev Entomol* **19**:385–405
65. Spieth HT (1968) **Evolutionary Implications of Sexual Behavior in *Drosophila*** In: Dobzhansky THEcht MKSteere WC, editors. *Evolutionary Biology* Boston, MA: Springer US pp. 157–193

66. Spieth HT (1966) **Drosophilid mating behavior: the behaviour of decapitated females** *Anim Behav* **14**:226–235
67. Spieth HT (1952) **Mating Behavior Within the Genus Drosophila (Diptera)** *Bulletin of the AMNH*
68. Sturtevant AH (1915) **Experiments on sex recognition and the problem of sexual selection in Drosophila** *J Exp Psychol Anim Behav Process* **5**:351–366
69. Swain B, von Philipsborn AC (2021) **Sound production in Drosophila melanogaster: Behaviour and neurobiology** *Advances in Insect Physiology* :141–187
70. Tompkins L, Hall JC, Hall LM (1980) **Courtship-stimulating volatile compounds from normal and mutant Drosophila** *J Insect Physiol* **26**:689–697
71. Wang L, Dankert H, Perona P, Anderson DJ (2008) **A common genetic target for environmental and heritable influences on aggressiveness in Drosophila** *Proc Natl Acad Sci U S A* **105**:5657–5663
72. Watanabe K, Chiu H, Pfeiffer BD, Wong AM, Hoopfer ED, Rubin GM, Anderson DJ (2017) **A Circuit Node that Integrates Convergent Input from Neuromodulatory and Social Behavior-Promoting Neurons to Control Aggression in Drosophila** *Neuron* **95**:1112–1128
73. Wohl MP, Liu J, Asahina K (2023) **Tachykininergic Neurons Modulate the Activity of Two Groups of Receptor-Expressing Neurons to Regulate Aggressive Tone** *J Neurosci* **43**:3394–3420
74. Yadav RSP, Ansari F, Bera N, Kent C, Agrawal P (2024) **Lessons from lonely flies: Molecular and neuronal mechanisms underlying social isolation** *Neurosci Biobehav Rev* **156**:105504
75. Zhang N, Guo L, Simpson JH (2020) **Spatial comparisons of mechanosensory information govern the grooming sequence in Drosophila** *Curr Biol* **30**:988–1001

Author information

R Sai Prathap Yadav[#]

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

[#]These authors contributed equally to this study.

Paulami Dey[#]

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

[#]These authors contributed equally to this study.

Faizah Ansari

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

Tanvi Kottat

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

P Pallavi Prabhu

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

Manohar Vasam

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

Shrinivas Ayyangar

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

Swathi Bhaskar S

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

Krishnananda Prabhu

Department of Biochemistry, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

Monalisa Ghosh*

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

*Present address: Gottfried Schatz Research Center, Medical University of Graz, Graz, Austria

Pavan Agrawal

Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi, India

ORCID iD: [0000-0002-4494-7859](https://orcid.org/0000-0002-4494-7859)

For correspondence: pavan.agrawal@manipal.edu

Editors

Reviewing Editor

Tihana Jovanic

Neuro-PSI, UMR-9197, CNRS, UPSaclay, Saclay, France

Senior Editor

Albert Cardona

University of Cambridge, Cambridge, United Kingdom

Reviewer #1 (Public review):

The study introduces an open-source, cost-effective method for automating the quantification of male social behaviors in *Drosophila melanogaster*. It combines machine-learning-based behavioral classifiers developed using JAABA (Janelia Automatic Animal Behavior Annotator) with inexpensive hardware constructed from off-the-shelf components. This approach addresses the limitations of existing methods, which often require expensive hardware and specialized setups. The authors demonstrate that their new "DANCE" classifiers accurately identify aggression (lunges) and courtship behaviors (wing extension, following, circling, attempted copulation, and copulation), closely matching manually annotated ground-truth data. Furthermore, DANCE classifiers outperform existing rule-based methods in accuracy. Finally, the study shows that DANCE classifiers perform as well when used with low-cost experimental hardware as with standard experimental setups across multiple paradigms, including RNAi knockdown of the neuropeptide Dsk and optogenetic silencing of dopaminergic neurons.

The authors make creative use of existing resources and technology to develop an inexpensive, flexible, and robust experimental tool for the quantitative analysis of *Drosophila* behavior. A key strength of this work is the thorough benchmarking of both the behavioral classifiers and the experimental hardware against existing methods. In particular, the direct comparison of their low-cost experimental system with established systems across different experimental paradigms is compelling. While JAABA-based classifiers have been previously used to analyze aggression and courtship (Tao et al., *J. Neurosci.*, 2024; Sten et al., *Cell*, 2023; Chiu et al., *Cell*, 2021; Isshi et al., *eLife*, 2020; Duistermars et al., *Neuron*, 2018), the demonstration that they work as well without expensive experimental hardware opens the door to more low-cost systems for quantitative behavior analysis.

Although the study provides a detailed evaluation of DANCE classifier performance, its conclusions would be strengthened by a more comprehensive analysis. The authors assess classifier accuracy using a bout-level comparison rather than a frame-level analysis, as employed in previous studies (Kabra et al., *Nat Methods*, 2013). They define a true positive as any instance where a DANCE-detected bout overlaps with a manually annotated ground-truth bout by at least one frame. This criterion may inflate true positive rates and underestimate false positives, particularly for longer-duration courtship behaviors. For example, a 15-second DANCE-classified wing extension bout that overlaps with ground truth for only one frame would still be considered a true positive. A frame-level analysis performance would help address this possibility.

In summary, this work provides a practical and accessible approach to quantifying *Drosophila* behavior, reducing the economic barriers to the study of the neural and molecular mechanisms underlying social behavior.

<https://doi.org/10.7554/eLife.105465.1.sa2>

Reviewer #2 (Public review):

Summary:

This manuscript addresses the development of a low-cost behavioural setup and standardised open-source high-performing classifiers for aggression and courtship behaviour. It does so by using readily available laboratory equipment and previously developed software packages. By comparing the performance of the setup and the classifiers to previously developed ones, this study shows the classifier's overperformance and the reliability of the low-cost setup in

recapitulating previously described effects of different manipulations on aggression and courtship.

Strengths:

The newly developed classifiers for lunges, wing extension, attempted copulation, copulation, following, and circling, perform better than available previously developed ones. The behavioural setup developed is low cost and reliably allows analysis of both aggression and courtship behaviour, validated through social experience manipulation (social isolation), gene knock (Dsk in Dilp2 neurons) and neuronal inactivation (dopaminergic neurons) known to affect courtship and aggression.

Weaknesses:

Aggression encompasses multiple defined behaviours, yet only lunges were analysed. Moreover, the CADABRA software to which DANCE was compared analyses further aggression behaviours, making their comparisons incomplete. In addition, though DANCE performs better than CADABRA and Divider in classifying lunges in the behavioural setup tested, it did not yield very high recall and F1 scores.

DANCE is of limited use for neuronal circuit-level enquiries, since mechanisms for intensity and temporally controlled optogenetic manipulations, which are nowadays possible with open-source software and low-cost hardware, were not embedded in its development.

<https://doi.org/10.7554/eLife.105465.1.sa1>

Reviewer #3 (Public review):

The preprint by Yadav et al. describes a new setup to quantify a number of aggression and mating behaviors in *Drosophila melanogaster*. The investigation of these behaviors requires the analysis of a large number of videos to identify each kind of behavior displayed by a fly. Several approaches to automatize this process have been published before, but each of them has its limitations. The authors set out to develop a new setup that includes very low-cost, easy-to-acquire hardware and open-source machine-learning classifiers to identify and quantify the behavior.

Strengths:

(1) The study demonstrates that their cheap, simple, and easy-to-obtain hardware works just as well as custom-made, specialized hardware for analyzing aggression and mating behavior. This enables the setup to be used in a wide range of settings, from research with limited resources to classroom teaching.

(2) The authors used previously published software to train new classifiers for detecting a range of behaviors related to aggression and mating and to make them freely available. The classifiers are very positively benchmarked against a manually acquired ground truth as well as existing algorithms.

(3) The study demonstrates the applicability of the setup (hardware and classifiers) to common methods in the field by confirming a number of expected phenotypes with their setup.

Weaknesses:

(1) When measuring the performance of the duration-based classifiers, the authors count any bout of behavior as true positive if it overlaps with a ground-truth positive for only 1 frame - despite the minimal duration of a bout is 10 frames, and most bouts are much longer. That

way, true positives could contain cases that are almost totally wrong as long there was an overlap of a single frame. For the mating behaviors that are classified in ongoing bouts, I think performance should be evaluated based on the % of correctly classified frames, not bouts.

(2) In the methods part, only one of the pre-existing algorithms (MateBook), is described. Given that the comparison with those algorithms is a so central part of the manuscript, each of them should be briefly explained and the settings used in this study should be described.

Taken together, this work can greatly facilitate research on aggression and mating in *Drosophila*. The combination of low-cost, off-the-shelf hardware and open-source, robust software enables researchers with very little funding or technical expertise to contribute to the scientific process and also allows large-scale experiments, for example in classroom teaching with many students, or for systematic screenings.

<https://doi.org/10.7554/eLife.105465.1.sa0>