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

R. Sai Prathap Yadav[#1], Paulami Dey[#1], Faizah Ansari[1], Tanvi Kottat[1], P. Pallavi Prabhu[1], Manohar Vasam[1], Shrinivas Ayyangar[1], Swathi Bhaskar S.[1], Krishnananda Prabhu[2], Monalisa Ghosh[1,3], Pavan Agrawal[1]*

1. Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi 576104, Karnataka, India.

2. Department of Biochemistry, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi 576104, Karnataka, India.

3. Present address: Gottfried Schatz Research Center, Medical University of Graz, Neue Stiftingtalstr. 6, 8010 Graz, Austria.

#These authors contributed equally to this study.

*Correspondence to: pavan.agrawal@manipal.edu

ORCID: 0000-0002-4494-7859

Abstract

Quantifying animal behavior is pivotal for identifying the underlying neuronal and genetic mechanisms involved. 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 *Drosophila* Aggression and Courtship Evaluator (DANCE), a low-cost, open-source platform that combines 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 address key limitations of rule-based algorithms, capturing dynamic behavioral variations more effectively. 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, which modulates aggression and courtship; RNAi-mediated downregulation of the neuropeptide Dsk; and optogenetic silencing of dopaminergic neurons, which promotes aggression. DANCE provides a cost-effective and portable solution for studying *Drosophila* behaviors in resource-limited settings or near natural habitats. Its accessibility and robust performance democratize behavioral neuroscience, enabling rapid screening of genes and neuronal circuits underlying complex social behaviors.

Keywords: *Drosophila*, aggression, courtship, machine learning, computer vision

Introduction

Detailed and accurate annotation and analysis of complex behaviors are necessary for understanding the underlying neural and molecular mechanisms. The fruit fly *Drosophila melanogaster* is one of the most accessible and well-studied model organisms for identifying the neuronal and molecular underpinnings of behavior. Multiple large-scale screens have been conducted in *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 the 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 distinct, stereotyped patterns. For example, 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 time-consuming and 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 programs, such as Ctrax, Caltech FlyTracker, and Deep Lab Cut (Branson et al., 2009; Eyjolfsdottir et al., 2014; Mathis et al., 2018), are widely used for tracking behaviors in *Drosophila*. Each comes with strengths and weaknesses. Ctrax (Branson et al., 2009) can accurately track fly position and movement, but identity switches remain a challenge, especially when tracking groups of flies. While both Ctrax and FlyTracker (Eyjolfsdottir et al., 2014) may produce identity switches, when groups of flies were tracked simultaneously, Ctrax led to inaccuracies that required manual correction using specialized algorithms such as FixTrax (Bentzur et al., 2021).

The effectiveness of various machine learning pipelines is eventually measured by comparing their output to human annotation, called 'ground-truthing'. Rule-based algorithm 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 used to quantify courtship; however, similar to CADABRA, it tends to miss true-positive events, leading to significant mis-scoring of behaviors under certain experimental conditions.

The 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 are utilized for training to encompass the dynamic variations in behaviors, allowing it to predict behaviors on the basis of learning from input data.

Several studies have developed JAABA-based behavioral classifiers for measuring aggression (Chiu et al., 2021; Chowdhury et al., 2021; Duistermars et al., 2018; Leng et al., 2020; Tao et al., 2024) and courtship (Gil-Martí et al., 2023; Pantalia et al., 2023). However, many of these studies did not make these classifiers publicly available (Duistermars et al., 2018; Gil-Martí et al., 2023; Hindmarsh Sten et al., 2025; 2020; Pantalia et al., 2023). In other cases, the reported approaches relied on specialized hardware, such as custom 3D printed parts (Chowdhury et al., 2021; Gil-Martí et al., 2023), or high-end machine-vision cameras (Chiu et al., 2021; Chowdhury et al., 2021; Duistermars et al., 2018; Hindmarsh Sten et al., 2025; Leng et al., 2020; Tao et al., 2024), limiting their accessibility and 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 setup built from off-the-shelf materials and consumer smartphones for behavioral recording. Compared with previous methods (Dankert et al., 2009; Ribeiro et al., 2018), the DANCE classifiers improved accuracy and reliability, while its low-cost hardware eliminates the need for specialized arenas and cameras. All classifiers and analysis codes are publicly available, enabling broad adoption, especially in resource-limited settings. Together, DANCE provides a powerful, accessible platform for behavioral screening and the discovery of 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 also designed a simple, low-cost recording setup constructed from repurposed transparent medicine blister packs, acrylic sheets, and paper tape, enabling easy behavioral recordings. To record these behaviors, we used Android smartphone cameras and an electronic tablet or smartphone serving 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 can be assembled from off-the-shelf components for less than USD 0.30 (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 that of the DANCE classifiers (Figure 1B-C).

To train the DANCE classifiers using JAABA (Kabra et al., 2013), for aggressive lunges, we used an 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 setup described in (Koemans et al., 2017) (Figure 1A; Figure 1— Figure Supplement 2). We tracked the position, motion, and interactions of pairs of flies across video frames using the Caltech FlyTracker (Eyjolfsdottir et al., 2014). To avoid data leakage, we randomly divided the acquired videos into two categories, "training videos" and “test videos,” to train and evaluate the 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 the DANCE classifiers against 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). Comparisons with manual ground-truth data revealed that the performance of the DANCE classifiers is comparable to that of human annotations and has higher sensitivity than rule-based algorithms (Figure 1D-E). The 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-truth and existing methods—CADABRA (Dankert et al., 2009) and the Divider assay classifier (Chowdhury et al., 2021). CADABRA tends to miss several true positive lunges, likely because of its rigid, rule-based framework, which cannot adapt to the dynamic variations in behavior. Figure 2A shows the lunge scores from 40 different videos using the ground-truth, the DANCE lunge classifier, CADABRA, and the Divider assay classifier. While the ground-truth and DANCE classifiers' outputs are comparable, CADABRA and the Divider assay classifier underscore lunges across

videos. We ground-truthed the DANCE lunge classifier against 40 ‘test videos’ (Figure 2A, Materials and Methods). Inter-observer validation confirmed that there were no significant differences between the two independent manual annotations (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-truth 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. hyperaggressive, >300 lunges (Figure 2E). DANCE scores remained comparable to ground-truth scores across all categories, whereas CADABRA and Divider underestimated the lunge counts (Figure 2B–E). Correlation analysis revealed a strong relationship between DANCE and ground-truth scores (Figure 2F, Supplementary File 2). In comparison, CADABRA and the Divider assay classifier showed a weaker correlation (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), whereas the Divider assay classifier, although also JAABA-based, was trained using data from a rectangular arena. Because JAABA classifiers rely on features influenced by the arena geometry, this mismatch likely reduced its accuracy in our circular setup. To further evaluate the performance, we computed the precision, recall, and F1 score (Figure 2I). The DANCE lunge classifier achieved a precision of 78.7%, recall of 73.1%, and an overall F1 score of 75.8%, exceeding the values obtained with other methods. Classifier robustness across multiple training videos, including the dataset used for inter-method comparisons (Video 9), is summarized in Figure 2—Figure Supplement 2.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.

DANCE classifiers to quantify courtship behaviors in *Drosophila*

The first report of *Drosophila* courtship behavior described stereotypic behaviors such as wing “scissor-like” movements, with males “swaying around the female,” licking, tapping, and mounting (Sturtevant, 1915). By the 2000s, studies revealed genes and neural circuits involved in courtship (Dickson, 2008; Pavlou and Goodwin, 2013). Automated analysis techniques for courtship exist, but their adoption has been limited by expensive hardware, reliance on custom parts, or a lack of publicly available code and classifiers (Duistermars et al., 2018; Gil-Martí et al., 2023; Koemans et al., 2017; Reza et al., 2013; Tao et al., 2024).

MateBook is a recent rule-based pipeline for automating the quantification of courtship behavior (Ribeiro et al., 2018). It relies on predefined rules derived from CADABRA (Dankert et al., 2009). To resolve ambiguities in the two overlapping flies when their trajectories are estimated from the video recordings, identities are assigned by relying on the distinct body sizes of the male and female flies, as the females are larger than the males. This size assumption is problematic in assays that use decapitated virgin females, which approximate male size. Such conditions are often used to assess male courtship independent of female behavioral feedback, for example, when evaluating pheromone effects (Cook and Cook, 1975; Spieth, 1966). In these contexts, rule-based approaches can introduce false positives and false negatives because the body-size criterion is not met.

To overcome these limitations, we trained and validated five new courtship classifiers using JAABA (Kabra et al., 2013). These classifiers quantify distinct stages of the male courtship ritual (Sokolowski, 2001), including wing extension, following, circling, attempted copulation, and copulation. To ensure robustness across conditions, training datasets

included videos with both decapitated and intact females (mated or virgin). We evaluated the DANCE classifier performance by comparing the outputs with the manual ground-truth and MateBook results. Because courtship behaviors differ in duration, we calculated a behavior index to enable direct comparisons between methods. Details of the classifier thresholds and criteria are provided in the Materials and Methods. To assess observer bias, annotations from two independent evaluators were compared, revealing no significant differences (Figure 2—Figure Supplement 1B–F).

Wing Extension

During unilateral wing extension, a 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 indices derived from manual ground-truth (gray) and the DANCE wing extension classifier (orange) across 15 videos with decapitated virgin females; the two measures are comparable, whereas MateBook (purple) systematically reports lower scores in most videos (Figure 3A). The DANCE wing extension index is strongly correlated with the ground-truth scores (Figure 3B-C) but weakly correlated with MateBook (Figure 3B, D). The classifier performance metrics confirm high reliability (precision 92.2%, recall 98.1%, F1 95.1%) (Figure 3E). Similar trends were observed in the mated female dataset (Figure 3—Figure Supplement 1).

Attempted copulation and copulation

Copulation typically lasts for approximately 15–25 minutes, and its duration is primarily determined by the male (MacBean and Parsons, 1967). Interrupted mating experiments have shown that sperm are transferred several minutes after copulation begins (Fowler, 1973; Tompkins et al., 1980). This distinction separates mounting into two outcomes—

successful copulation and unsuccessful attempted copulation—for which we developed separate classifiers (see Materials and Methods).

The DANCE attempted copulation classifier closely matched the ground-truth across videos (Figure 4A–B). Compared with MateBook, which often overestimates attempted copulation events, the DANCE classifier provides more consistent detection. The correlation with the ground-truth was stronger for DANCE as compared to MateBook (Figure 4C–D), and the performance metrics for DANCE were robust (precision 82.6%, recall 89.2%, F1 85.8%; Figure 4E).

For copulation, which is trained on videos using mated and decapitated virgin females, the DANCE copulation classifier also matches the ground-truth with near-perfect performance (Figure 4—Figure Supplement 1). MateBook performs reasonably well for copulation, but alignment or arena-detection errors in some recordings cause occasional false negatives or positives (Figure 4—Figure Supplement 1A, videos #13 and #18).

Circling

A male circling around a female is a distinct courtship element implicated in female re-stimulation (Kessler, 1962). Variations in circling frequency across species contribute to reproductive isolation (Brown, 1965).

We evaluated the DANCE circling classifier on both the decapitated-virgin and mated-female datasets (Figure 5 and Figure 5—Figure Supplement 1). The circling indices from ground-truth and DANCE were comparable across videos (Figure 5A–B), whereas MateBook often under-represented circling in individual recordings (Figure 5A). The DANCE circling index correlated strongly with ground-truth scores as compared to MateBook (Figure 5C-D), and classifier metrics indicated robust performance (precision 98.0%, recall 92.1%, F1 95.0%; Figure 5E).

Following

During following, the male tracks the female's movement to initiate subsequent courtship acts (Spieth, 1968). Because following is a relatively continuous and readily defined behavior, both the MateBook and the DANCE following classifier performed well (Figure 5—Figure Supplement 2A–D). However, the DANCE following classifier produced more balanced scores (precision 91.2%, recall 91.1%, F1 91.1%) compared with MateBook (precision 65.8%, recall 83.4%, F1 73.5%; Figure 5—Figure Supplement 2E), indicating lower rates of false positives and false negatives for DANCE.

Finally, we performed frame-level analyses in addition to bout-level evaluations to provide a more granular assessment of the courtship classifiers (see Materials and Methods). Frame-level metrics showed only marginal reductions in performance compared with bout-level metrics (Figure 5—Figure Supplement 3). Together, these results demonstrate that the DANCE classifiers provide a reliable and accurate means to quantify both aggression and courtship behaviors, supporting subsequent benchmarking using the DANCE hardware.

DANCE hardware

Existing setups for recording *Drosophila* aggression and courtship (Figure 1A; Figure 1—Figure Supplement 1–2) present several practical challenges that limit their broad adoption. These include the need for complex, custom-fabricated components, 3D-printed parts, specialized machine-vision cameras and backlights, and considerable technical expertise for data acquisition and processing (Chowdhury et al., 2021; Dankert et al., 2009; Gil-Martí et al., 2023; Koemans et al., 2017). Setting up some aggression assays (Dankert et al., 2009; Dierick, 2007) also requires coating chambers with fluon to prevent flies from walking on the walls, which is labor-intensive.

To provide a low-cost, easy-to-assemble alternative, we developed the DANCE hardware (Figure 6), an inexpensive, scalable, and robust system for recording *Drosophila* aggression and courtship behaviors.

The DANCE hardware consists of readily available off-the-shelf components, including transparent medicine blister packs (tablet foils) used as recording chambers, which are mounted on 2 mm acrylic base plates and secured with paper tape (Figure 6A–D, Supplementary Videos 1–2). Instead of using machine-vision cameras, DANCE employs widely available Android smartphones for recording and substitutes backlights with tablets or smartphones displaying a white screen to provide uniform illumination (Figure 6E–G). Aggression and courtship behaviors were recorded at 30 fps and 1080p resolution.

For the aggression assays (Figure 6A–B, E), the blister foil was slid over a base plate containing an apple-juice agar food layer, which served as the interaction arena (Supplementary Videos 1–2). For courtship assays, the blister foil was bisected and fitted with a thin X-ray-film separator comb that kept males and females apart until the start of recording, when the comb was removed to allow interaction (Supplementary Videos 3–4). Because the tablet or smartphone screens used as the backlight generate heat, we placed a transparent acrylic spacer above the backlight to create a 4 mm air gap for heat dissipation (Figure 6E–G; Supplementary Videos 2, 4). This modification was essential for maintaining consistent behavioral recordings (Figure 7).

The DANCE hardware is intentionally modular, allowing users to adapt the setup to their experimental needs. The detailed assembly, cleaning, and reuse protocols are provided in the Materials and Methods and on the GitHub page of the project.

Benchmarking DANCE hardware

We benchmarked the DANCE hardware by applying the validated DANCE classifiers to videos recorded in the DANCE setup and comparing the results with those from established recording systems. Wild-type males displayed quantitatively similar levels of courtship and aggression in DANCE arenas and in pre-existing setups (Figure 7; Supplementary Video 5).

To test whether DANCE reproduces established behavioral findings, we examined the effects of social isolation and enrichment on aggression and courtship. Previous studies have shown that single-housing (SH) 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 al., 2024). We found that both the DANCE and pre-existing setups captured similar and statistically significant differences in courtship behaviors between SH and group-housed (GH) males (Figure 7C–J). These results confirm that DANCE reliably detects behavioral modulation by social experience.

We next compared an established aggression assay (Dankert et al., 2009); Figure 7K) with the DANCE aggression setup (Figure 7L). Both systems detected aggressive lunges in SH flies and showed consistent differences between SH and GH conditions (Figure 7M–N). Thus, DANCE hardware provides comparable sensitivity to conventional machine-vision-based setups while being more accessible.

We also tested whether diet composition alters aggression in DANCE assays, as nutrient availability and microbiome interactions can influence male aggression (Jia et al., 2021; Lim et al., 2014). Replacing yeast granules in the diet with yeast extract powder reduced baseline aggression (Figure 7O). These findings demonstrate that the DANCE hardware is sensitive enough to detect diet-dependent behavioral differences.

To test whether DANCE is compatible with neurogenetic manipulations, we used RNAi-mediated knockdown of the neuropeptide Drosulfakinin (Dsk) in insulin-producing neurons using the *dilp2*-GAL4 driver. Consistent with our previous findings (Agrawal et al., 2020), SH males with Dsk knockdown exhibited significantly increased aggressive lunges compared to controls (Figure 7P–Q).

We then evaluated DANCE's suitability for optogenetic assays, a common approach to dissect neural circuits underlying aggression (Hoopfer et al., 2015; Wohl et al., 2023; Yadav et al., 2024). An earlier study has shown that constitutive silencing of broad populations of dopaminergic neurons using *TH*-GAL4 produces unhealthy flies with impaired locomotion that rarely fight (Alekseyenko et al., 2013). We reasoned that transient, light-controlled silencing could overcome these limitations and tested this using an optogenetic module integrated with DANCE (Figure 7—Figure Supplement 1A–B; DANCE GitHub).

Optogenetic silencing of dopaminergic neurons expressing the green-light-sensitive anion channelrhodopsin GtACR1 (Govorunova et al., 2015; Mohammad et al., 2017) during 20-minute interactions resulted in a significant increase in aggressive lunges in SH flies (Figure 7R). In addition, we observed higher frequencies of wing flicks and high-intensity aggressive behaviors such as boxing and tussling (Supplementary Video 6). Importantly, continuous silencing for 12 hours did not alter general locomotor activity between the GH and SH flies (Figure 7—Figure Supplement 2), confirming that these effects were not due to impaired movement.

We also examined whether arena size influenced behavior. *TH*-GAL4 > *UAS-GtACR1* males showed increased lunging versus controls across all sizes tested (13, 17 and 21 mm; Figure 7—Figure Supplement 3A–C), indicating the optogenetic effect was robust to chamber dimensions. In courtship assays (11, 13 and 17 mm arenas; Figure 7—Figure

Supplement 4A–L), single-housed males exhibited more wing extension, attempted copulation and circling than group-housed males, while following was similar. Although statistical tests were limited to within-size comparisons, the distributions suggest a modest decrease in interaction frequency in larger arenas, consistent with previous observations that larger chambers reduce encounter rates (Chowdhury et al., 2021).

Taken together, these results show that DANCE hardware provides reliable, reproducible behavioral measurements and is compatible with genetic and optogenetic manipulations as well as environmental perturbations.

Discussion

Here, we present the DANCE assay, an easy-to-use, modular, and robust analysis pipeline with inexpensive hardware to record and quantify aggression and courtship behaviors. We developed six novel behavioral classifiers using supervised machine learning to accurately quantify the aggression and courtship behaviors of *Drosophila* males. The hardware component of DANCE, fabricated from repurposed and low-cost materials, provides a practical alternative to conventional machine-vision setups. The DANCE setup can be built for less than 0.30 USD, representing an approximately 10,000-fold reduction in cost compared with standard systems. Despite this simplicity, its performance was comparable to that of more specialized and expensive hardware, validating DANCE as a reliable and accessible behavioral platform. This accessibility enables rapid behavioral screening and wider adoption by the neuroscience community, including resource-limited laboratories and teaching environments.

Various components of the DANCE assay, such as the behavioral classifiers, hardware design, and analysis code, are publicly available and can be used independently. This open, modular design gives researchers the flexibility to customize classifiers for specific

behavioral paradigms or incorporate new data without developing a classifier from scratch. Although not implemented here, DANCE can be readily extended to real-time feedback experiments using open-source reactive programming tools such as Bonsai (Lopes et al., 2015). This framework can also support the development of future classifiers for additional social behaviors, including aggressive acts such as fencing, wing flicking, tussling, chasing, or female headbutting (Chen et al., 2002; Nilsen et al., 2004), and courtship behaviors such as male tapping, licking, or female rejection (Sokolowski, 2001).

Such high-resolution analysis of complex social interactions can provide deeper understanding of mating dynamics, sexual selection, and the influence of genetics and evolution. Further, quantifying behavioral dynamics can reveal the temporal organization of individual components of behavior (Nilsen et al., 2004; Seeds et al., 2014; Simon and Heberlein, 2020; Zhang et al., 2020). DANCE can also serve as a flexible framework for studying complex behaviors across multiple *Drosophila* species and other insects, enabling comparative and evolutionary analyses. The adaptability and portability of the DANCE assay make it particularly useful for ethologists examining insect behavior in semi-natural or field-like environments. Together, these features position DANCE as a bridge between laboratory-based ethology and ecological studies of natural behavior.

Materials and Methods

The details of all the custom codes, analysis pipelines, sample files used 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 12 hr:12 hr light□dark cycle. All the assays were performed at 25°C with 65% relative humidity, unless mentioned otherwise. For the aggression and courtship experiments, Canton-S male flies were collected within 24 hours of eclosion and housed in groups (20 male flies per vial, 90 mm in length and 25 mm in diameter) or isolated (1 male fly per vial, 70 mm in length and 10 mm in diameter) for 6 days.

The following fly lines were acquired from the Bloomington Drosophila Stock Center (BDSC), USA: *TH*-GAL4 (RRID:BDSC_51982), *Dilp2*-GAL4 (RRID:BDSC_37516), *Dsk*-RNAi (RRID:BDSC_25869) and attP2 empty vector control (RRID:BDSC_36303). *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

Aggression assays were 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 2 mm 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 with screws. Fluon (Insect-a-slip, Bioquip: cat #2871B**)** was applied to the arena walls, which were left to dry overnight to create a slippery surface and prevent climbing. Sigmacote (Sigma□Aldrich: SL2) was used to coat the sliding lid to reduce walking on the arena ceiling. The chamber was placed on a food plate containing commercial apple juice (without added sugars), 2.5% w/v sucrose (HiMedia: GRM601), and 2.25% w/v agar (HiMedia: GRM026). For experiments to test the 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 the fly food. For optogenetic experiments, LEDs emitting 520–540 nm light (Lumileds: 2835) were used at an intensity of 0.0004 µW, which was measured with a power meter (Newport: 843-R). For all the aggression assays, the flies were allowed to acclimatize in the arena for 5 minutes, after which the activity was recorded for 20 minutes. The assays were performed during ZT0 - ZT2.5, i.e., during the first 2.5 hours of the morning activity peak.

Courtship assay

Courtship assays were performed as described previously (Koemans et al., 2017; Ribeiro et al., 2018). Single pairs of males and females were introduced into individual arenas of an 18-well courtship chamber (10 mm diameter). Male and female flies were introduced into one half of the chamber by 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 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 mated females, 20 females were housed with 10 males for 4–6 days. For decapitated virgin females, 2–4-day-old virgins were anesthetized with CO_2 and decapitated immediately before the assay.

DANCE hardware

Circular transparent medicine blister packs serve as aggression or courtship arenas. Blister packs were mounted on 2 mm acrylic base plates and secured with paper tape. The arena dimensions were as follows: aggression, 13 mm × 5.5 mm (diameter × height) (also tested: 17 × 4 mm and 21 × 8 mm); courtship, 11 mm × 4.5 mm (also tested: 13 × 5.5 mm and 17 × 4 mm). Arena walls and roofs were coated with Sigmacote (Sigma□Aldrich: SL2) using cotton swabs (Solimo, Amazon India) to prevent flies from climbing.

A thin 2 mm acrylic base plate carrying the food layer (2.25% w/v agar in commercial apple juice with 2.5% w/v sucrose) and a side spacer were assembled and held together with paper tape (Supplementary Video 1). The tip of the food plate was covered with paper tape to allow smooth sliding of the blister foil without damaging the food surface. Flies were introduced through loading holes (2 mm diameter), and paper tape strips were used to prevent food damage during assembly and to seal small gaps to avoid escape (Supplementary Video 2).

Arenas were reusable for approximately 30–40 times when washed in 0.05% Alcojet (Alconox: 1401-1) and air-dried. Heat exposure during cleaning was avoided to prevent deformation.

The standard DANCE courtship arena used transparent medicine blister packs measuring 11 mm in diameter and 4.5 mm in height. Additional arena sizes of 13 × 5.5 mm and 17 × 4 mm (diameter × height) were also tested. The top surface of each arena strip (five wells in a row) was modified by cutting thin slits with a sharp razor to insert a separator comb made from repurposed X-ray film. The foil and 2 mm acrylic base plate, which contained a loading hole (2 mm diameter), were joined with paper tape. The courtship base plate contained two loading holes so that male and female flies could be introduced on either side of the separator comb. The assembly was sealed at both ends with paper tape to prevent gaps from escaping (Supplementary Video 3).

Before recording, the separator comb was gently lifted to allow interaction between the flies, taking care to keep the slit edges intact and avoid distortion during filming (Supplementary Video 4). Blister packs modified with slits and separator combs are typically reusable for ~10–15 experiments, after which the slits tend to widen, and new blister packs are recommended.

Video acquisition

For the traditional setup, the interaction of the flies was recorded using machine-vision cameras (DMK 33UX252 USB 3.0 monochrome camera). 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.

For DANCE hardware testing, various Android smartphone cameras were used (Huawei Y9 2019; OnePlus Nord CE 2 Lite 5G, model: CPH2381; Redmi Note 11 Pro+ 5G, model: 221116SI) at 30 fps, 1080p resolution in H.264 (.mp4) format for 15 minutes (courtship) or 20 minutes (aggression). An electronic tablet (iPad Air, 5th Generation) or smartphone (iPhone 13) running a 'white screen light app' served as the background illumination source. The screen brightness was adjusted within the app to optimize contrast for fly tracking. 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 prevent the DANCE arenas from becoming hot. Devices were recommended to be placed in airplane mode during recordings to avoid interruptions.

Tracking flies using FlyTracker

Fly locations, body orientations, and interactions were tracked using Caltech FlyTracker (Eyjolfsdottir et al., 2014). These data were then pushed to the JAABA pipeline to develop DANCE classifiers. Occurrences of identity switches were corrected using the FlyTracker "visualizer" identity-correction tool. Tracking accuracy and identity swap quantification were validated by a semi-manual inspection of flagged frames and intervals (see Supplementary File 5 for details).

Pre-existing algorithms used for benchmarking

CADABRA

CADABRA (Dankert et al., 2009) analyzes two-fly interactions based on spatial and postural features, classifying lunges, wing threats, circling, wing extension, and copulation based on fixed rules. The specific CADABRA definitions are described in the subsequent section.

Divider Assay

The Divider Assay (Chowdhury et al., 2021) uses a 3D-printed rectangular chamber with 12 arenas (13 mm × 4.5 mm, W × H), each separated by an opaque divider. Behavioral data are analyzed with a custom FlyTracker-JAABA pipeline. Since the Divider assay classifier was trained on recordings from a rectangular geometry, it can affect transferability to circular arenas.

MateBook

MateBook (Ribeiro et al., 2018); https://github.com/Dicksonlab/MateBook) uses machine vision to track flies and classify male courtship behaviors (following, wing extension, orientation, copulation, and circling). The outputs include a .tsv file with bout statistics and an ethogram. For comparison with the DANCE classifiers, the MateBook persistence filters were adjusted so that the minimum bout duration threshold was 0.33 s (10 frames at 30 fps) for all behaviors except copulation, which retained a 45 s threshold. This adjustment ensured comparable persistence criteria between the MateBook and DANCE analyses.

Developing DANCE classifiers

JAABA (Kabra et al., 2013) was used to train classifiers iteratively: true bouts were labeled, obvious non-bouts assigned as “None,” and false positives were relabeled until performance plateaued.

The DANCE lunge classifier was trained on 11 independent videos with classifier accuracy improving progressively from Video 1 to Video 9. Peak performance was achieved after inclusion of the ninth training video**,** which provided the best balance of precision, recall, and F1 score. Adding further data (Videos 10 and 11) did not enhance classifier accuracy and instead produced a slight reduction in precision and recall, likely due to increased behavioral variability across sessions. Therefore, the final lunge classifier was trained on nine videos, which yielded the most robust and generalizable model.

Courtship classifiers were developed following the same iterative procedure. Training sets included videos with both decapitated and intact females (mated or virgin) to capture behavioral variability and ensure robust generalization. Independent classifiers were trained for wing extension, following, circling, attempted copulation, and copulation. All classifiers were validated using manually annotated “ground-truth” test videos that were not included in the training set.

Courtship training sets included decapitated virgin videos to enrich attempted copulation and circling bouts and to balance “None” class examples for following and copulation. Wing extension used mated females only.

The training data volumes (frames and approximate durations) were as follows:

Behavior	**Frames**	**Approx. duration (s)**
Lunge	1449	~48

Wing extension	99,947	~3,332
Attempted copulation	39,513	~1,317
Copulation	56,979	~1,899
Circling	14,396	~480
Following	25,787	~860

Test videos were manually annotated in JAABA ground-truthing mode before any classifier predictions were examined, and annotations were performed independently of classifier outputs.

Manual behavioral annotations and inter-annotator reliability

Manual annotations ("ground-truth") were generated using JAABA's ground-truthing mode by labeling behavioral bouts frame-by-frame. To assess observer bias, subsets of videos were annotated independently by two evaluators and compared using non-parametric tests. Where relevant, the results describe inter-annotator comparisons.

Characterization of male aggression and courtship behaviors

The DANCE classifiers were trained and validated in JAABA using established behavioral definitions from previous studies (Dankert et al., 2009; Ribeiro et al., 2018), with input from experienced users.

Behavior definitions

Lunge: As defined by (Dankert et al., 2009; Ribeiro et al., 2018), "the attacking fly rises on hind legs, lifting its long body axis by 45°, then snaps down on its opponent's body with its head at ~200 mm/s."

Wing extension: The angle between the body axis and wing tip line exceeds 30°,

persisting for at least 13 frames (0.5 s at 25 fps) (Dankert et al., 2009; Ribeiro et al., 2018).

Attempted copulation: Abdominal curling without mounting, or mounting lasting ≥0.33 s but <45 s (10–1350 frames at 30 fps).

Copulation: Mounting lasting ≥45 s (>1350 frames at 30 fps), adapted from (Dankert et al., 2009; Ribeiro et al., 2018).

Circling: Sideways drift around the female in a circular path at constant velocity, persisting for ≥13 frames (Dankert et al., 2009; Ribeiro et al., 2018).

Following: The male remains 2–5 mm behind the female while both walk at ≥2 mm/s, persisting for ≥25 frames (Dankert et al., 2009; Ribeiro et al., 2018).

To account for variability in female mating condition and body size, training datasets included videos of males paired with decapitated virgin, intact virgin, and mated females. For the copulation classifier, videos in which males were paired with virgin females to capture prolonged occlusion events characteristic of mating were used for training.

For duration-based classifiers, a post-processing filter was applied to exclude bouts shorter than 98% of those observed in manual annotations, ensuring consistency with human-defined behavioral durations.

This framework can be readily adapted by the research community to develop additional behavioral classifiers. Owing to file size limitations, training videos are not hosted online but are available upon request.

Manual behavioral annotations

To quantitatively evaluate classifier performance, manual behavioral annotations ("ground-truth") were generated using JAABA's ground-truthing mode (Kabra et al., 2013). Classifier robustness was assessed on unseen videos comprising the testing set. Each testing video was first manually annotated by identifying behavioral bouts as "true behavior," independent of the classifier's output. These same videos were then processed through the trained classifier using JAABAPlot, and the results were compared as described below.

Comparison of manual and DANCE annotations

The classifier outputs were compared with manually annotated ground-truth data at both the bout and frame levels, depending on the classifier type. For the single-frame lunge classifier, comparisons were based on total bout counts, whereas for all duration-based courtship classifiers (wing extension, following, circling, attempted copulation, and copulation), comparisons were made using bout durations and frame level annotations.

For each assay, a behavioral index was calculated as the proportion of frames in which the male engaged in the specified behavior. This was obtained by dividing the total number of frames annotated for that behavior by the total number of frames in the recording. Regression analyses and performance metrics were computed using either bout counts or behavioral indices, depending on the classifier type, in GraphPad Prism 8 (GraphPad Software). Manual annotations and classifier outputs were compared to identify true positives (TP), false positives (FP), and false negatives (FN), at either frame-level or bout-level.

Bout-level analysis

A predicted bout was scored as a TP if it overlapped with a ground-truth bout by at least one frame (~33 ms at 30 fps), consistent with previous studies (Leng et al., 2020). When

multiple predicted bouts overlapped a single ground-truth bout, they were collectively counted as one TP. Conversely, when a single predicted bout overlapped multiple ground-truth bouts, the TP count equaled the number of ground-truth bouts. Predicted bouts with no overlap were scored as FP, and ground-truth bouts with no overlapping prediction were scored as FN. These same criteria were applied to both bout-level and frame-level evaluations, with the latter accounting for the total number of frames contributing to TP, FP, and FN classifications to provide a more granular measure of accuracy (see Figure 5—Figure Supplement 3).

Frame-level analysis

A predicted frame was scored TP if it matched the ground-truth frame; frames predicted as behavior only by the classifier were FP; frames annotated as behavior only by ground-truth were FN. These frame-level TP, FP, FN were then similarly used to calculate precision, recall, and F1 score, providing a more granular measure of classifier accuracy.

Performance metrics

A custom Python script was used to calculate overlaps and derive standard classification metrics. The precision, recall, and F1 scores were computed using the following formulas:

$$precision = \frac{(number\ of\ true\ positives)}{(number\ of\ true\ positives) + (number\ of\ false\ positives)}$$

$$recall = \frac{(number\ of\ true\ positives)}{(number\ of\ true\ positives) + (number\ of\ false\ negatives)}$$

$$F1\ score = 2\ x\ \frac{(precision\ x\ recall)}{(precision +\ recall)}$$

Precision represents the fraction of correctly predicted positive observations among all predicted positives, recall represents the fraction of correctly predicted positive observations among all actual positives, and the F1 score provides the harmonic mean of

precision and recall. These metrics were used to quantify and compare the performance of the DANCE classifiers against manually annotated ground-truth and existing algorithms.

For duration-based behaviors, the behavioral index was used as a continuous variable in regression and performance analyses. For the lunge classifier, which identifies discrete one-frame events, comparisons were made using total bout counts.

Statistical analysis

All the statistical analyses were performed using *GraphPad Prism 8*, custom Python scripts, or *Microsoft Excel*. For non-normally distributed data, non-parametric tests such as the Mann–Whitney U test or Kruskal–Wallis ANOVA with appropriate post hoc corrections were used.

Supplementary Video legends:

1. 3D rendered DANCE Aggression hardware.
2. Using the DANCE hardware setup for recording aggression.
3. 3D rendered DANCE Courtship hardware
4. Using the DANCE hardware setup for recording courtship.
5. Aggression and courtship behaviors recorded in DANCE hardware.
6. Optogenetic silencing of dopaminergic neurons in DANCE shows increased aggression.

Supplementary File descriptions:

1. Bill of materials for the DANCE setup and comparison with existing setups.
2. Comparison between the DANCE lunge classifier, ground-truth, and existing methods.
3. Definitions of the behavioral classifiers.

4. Comparison between the DANCE courtship classifiers, ground-truth, and MateBook.
5. Quantifying identity swaps to validate tracking across setups.

Acknowledgment

We thank Barry Dickson (Queensland Brain Institute, Australia) for insightful discussions and suggestions related to MateBook and Ben Arthur (HHMI, Janelia Research Campus, USA) for guidance in setting up the MateBook analysis. We are grateful to Mayank Kabra, Kristin Branson, and colleagues for developing JAABA and for their ongoing support to the user community. We thank Ulrike Heberlein (HHMI, Janelia Research Campus, USA), Gaurav Das (NCCS, Pune), and the Bloomington Drosophila Stock Center (NIH P40OD018537) for providing fly stocks. We acknowledge Santhosh Chidangil (MAHE) for assisting with the LED power measurements and Santosh D'Mello (LSU Shreveport) for helpful discussions. We are also grateful to Gaurav Das and Toshiharu Ichinose (Tohoku University, Japan) for critical reading and feedback on the manuscript. Figures were created with BioRender.com.

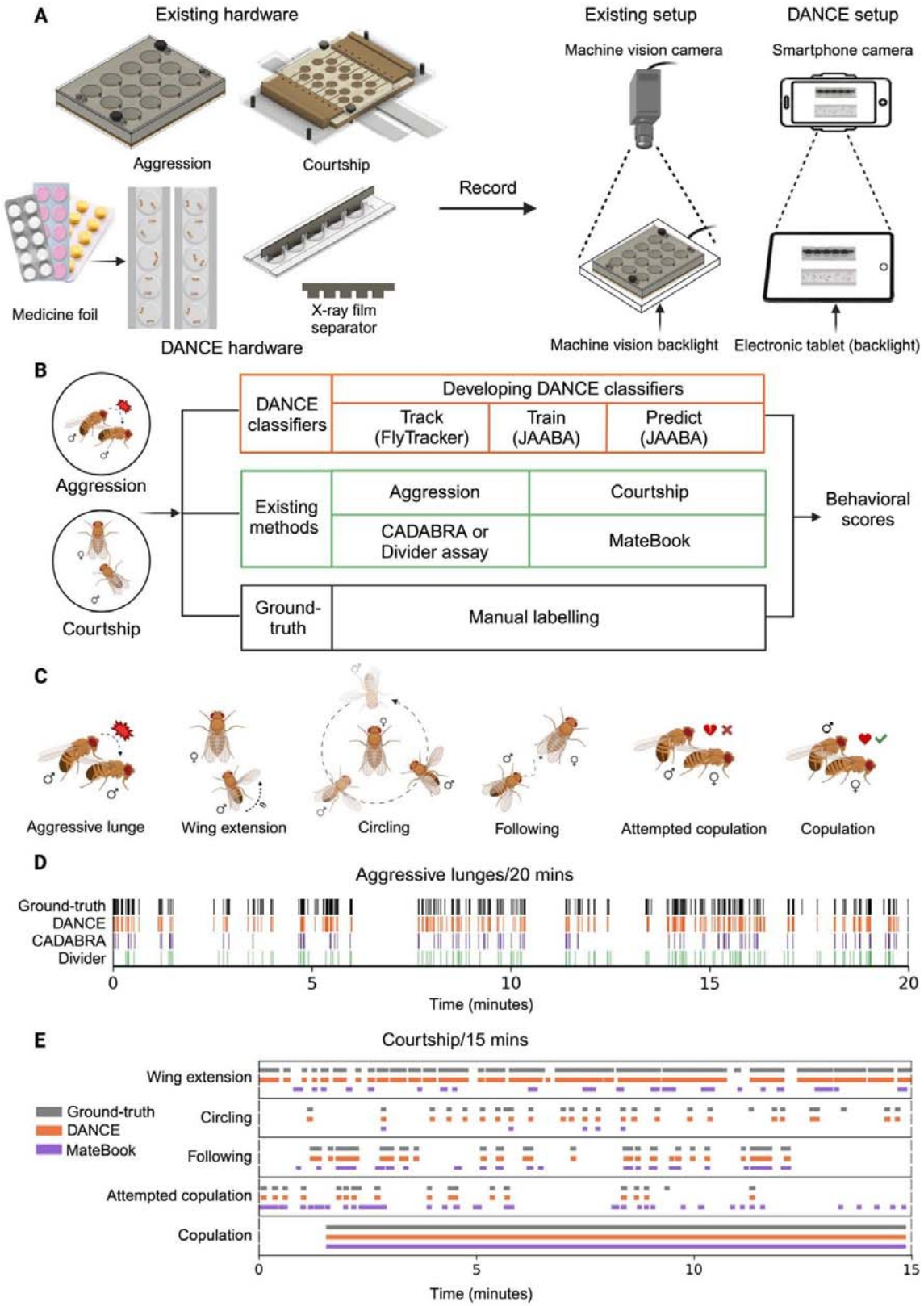
Funding

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).

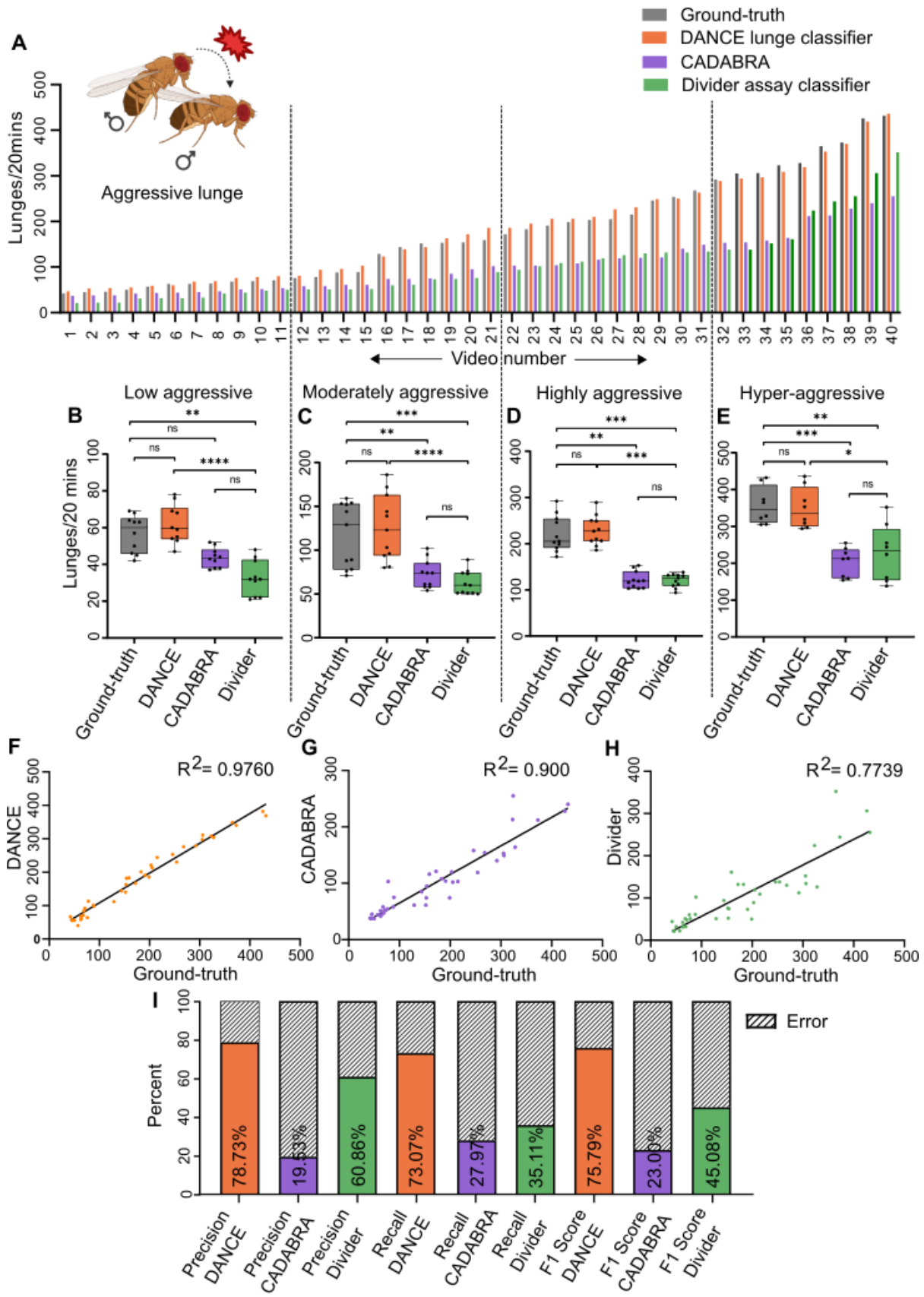
Competing interests: RSPY, FA and PA are listed as inventors on a published patent related to this work by the Indian patent office, titled ‘Device For Measuring Complex Social Behaviors In Small Insects’, Application No. 202441072884.

Figures and Figure Legends


A
Existing hardware
Aggression
Courtship
Medicine foil
X-ray film separator
DANCE hardware
Record
Existing setup
Machine vision camera
Machine vision backlight
DANCE setup
Smartphone camera
Electronic tablet (backlight)
B
Aggression
Courtship
DANCE classifiers
Developing DANCE classifiers
Track (FlyTracker)
Train (JAABA)
Predict (JAABA)
Existing methods
Aggression
Courtship
CADABRA or Divider assay
MateBook
Ground-truth
Manual labelling
Behavioral scores
C
Aggressive lunge
Wing extension
Circling
Following
Attempted copulation
Copulation
D
Aggressive lunges/20 mins
Ground-truth
DANCE
CADABRA
Divider
0
5
10
15
20
Time (minutes)
E
Courtship/15 mins
Wing extension
Circling
Following
Attempted copulation
Copulation
Ground-truth
DANCE
MateBook
0
5
10
15
Time (minutes)

681

Figure 1. The DANCE assay provides an accessible approach for quantifying aggression and courtship behaviors. (A) Comparison of existing machine-vision camera hardware with the simplified, low-cost DANCE hardware for behavior acquisition. (B) Workflow for developing DANCE classifiers, including training, benchmarking against existing methods and manual ground-truthing to generate behavioral scores. (C) Behavioral classifiers developed to quantify male aggression (lunge) and courtship (wing extension, circling, following, attempted copulation, and copulation). (D) Representative raster plots comparing ground-truth, DANCE, CADABRA, and Divider assay performance for aggression. (E) Representative raster plots comparing ground-truth, DANCE, and MateBook performance for courtship.


A
Ground-truth
DANCE lunge classifier
CADABRA
Divider assay classifier
Aggressive lunge
Lunges/20mins
Video number
Low aggressive
Moderately aggressive
Highly aggressive
Hyper-aggressive
B
C
D
E
Lunges/20 mins
Ground-truth
DANCE
CADABRA
Divider
F
G
H
R2= 0.9760
R2= 0.900
R2= 0.7739
DANCE
CADABRA
Divider
Ground-truth
I
Percent
Error
78.73%
19.53%
60.86%
73.07%
27.97%
35.11%
75.79%
23.03%
45.08%
Precision DANCE
Precision CADABRA
Precision Divider
Recall DANCE
Recall CADABRA
Recall Divider
F1 Score DANCE
F1 Score CADABRA
F1 Score Divider

692

Figure 2. Validation of the DANCE lunge classifier for quantifying male aggression. (A) Lunge scores from 20-minute-long videos obtained by the ground-truth (gray), the DANCE lunge classifier (orange), CADABRA (purple) and Divider assay classifier (green). (B-E) Comparison of lunge scores across different aggression levels, based on manual scoring and predictions from DANCE, CADABRA, and Divider: (B) 0–70 lunges 'low aggressive' (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 the DANCE 'lunge classifier' *vs.* manual scores (R^2=0.9760, n=40). (G) Regression of the CADABRA *vs.* the DANCE lunge classifier (R^2=0.9, n=40). (H) Regression of the Divider assay lunge classifier score *vs.* manual score (R^2=0.7739, n=40). (I) Precision, recall, and F1 scores of the DANCE lunge classifier compared with those of CADABRA and Divider.

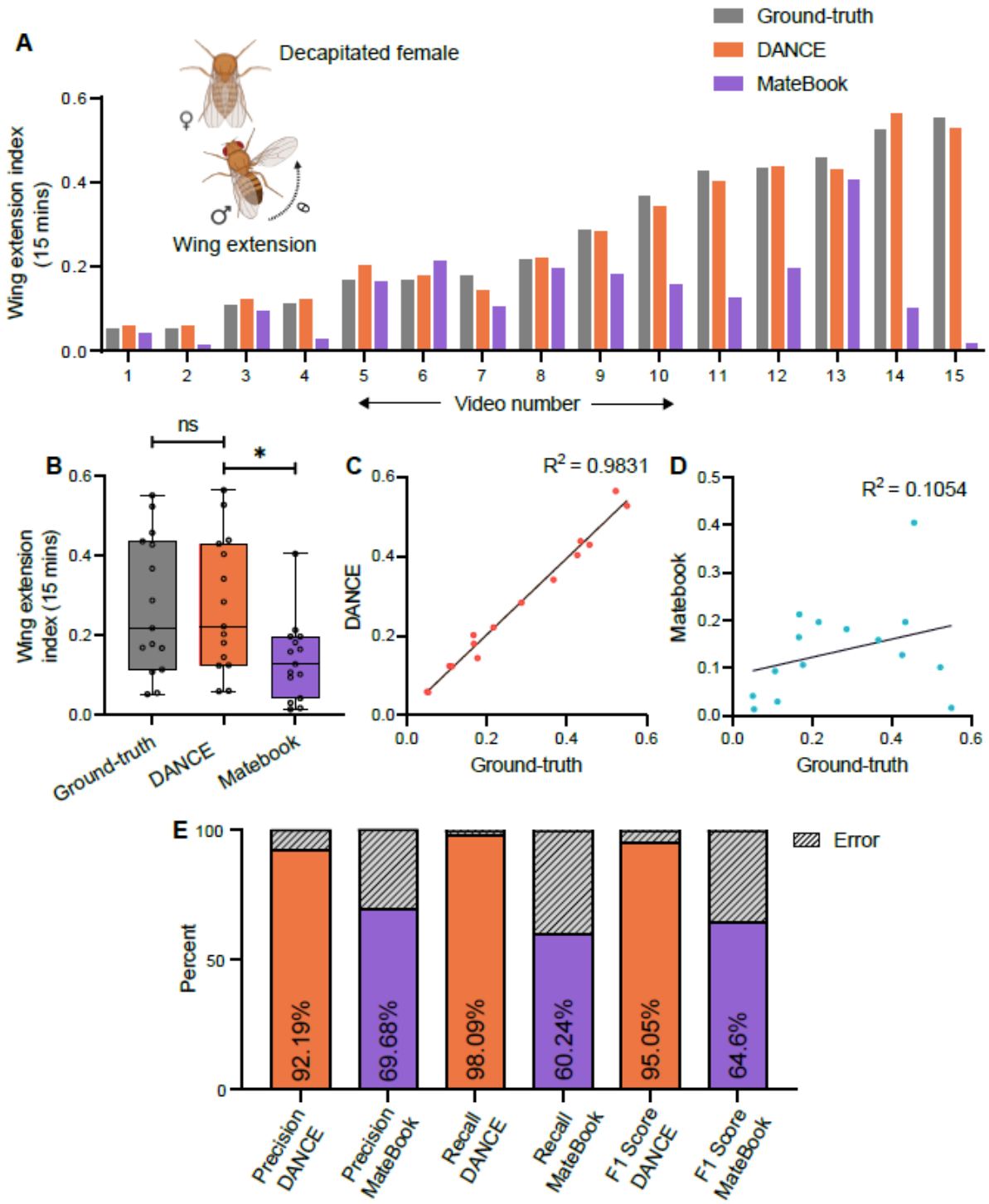


Figure 3. Evaluation of the DANCE wing extension classifier for quantifying courtship behavior. (A) Wing extension index of males from 15 minutes long videos scored by manual ground-truth (gray), the DANCE wing extension classifier (orange) and MateBook (purple), with decapitated virgin females. MateBook underscored wing

extension across multiple videos (Friedman's ANOVA with Dunn's test: ground-truth *vs* MateBook p=0.0020, n=15; ground-truth *vs* DANCE p>0.9999). (B) Comparison of ground-truth, DANCE, and MateBook wing extension scores (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) Regression of MateBook *vs.* ground-truth (R^2=0.1054, n=15). (E) Precision, recall, and F1 score of the DANCE wing extension classifier and MateBook relative ground-truth scores.

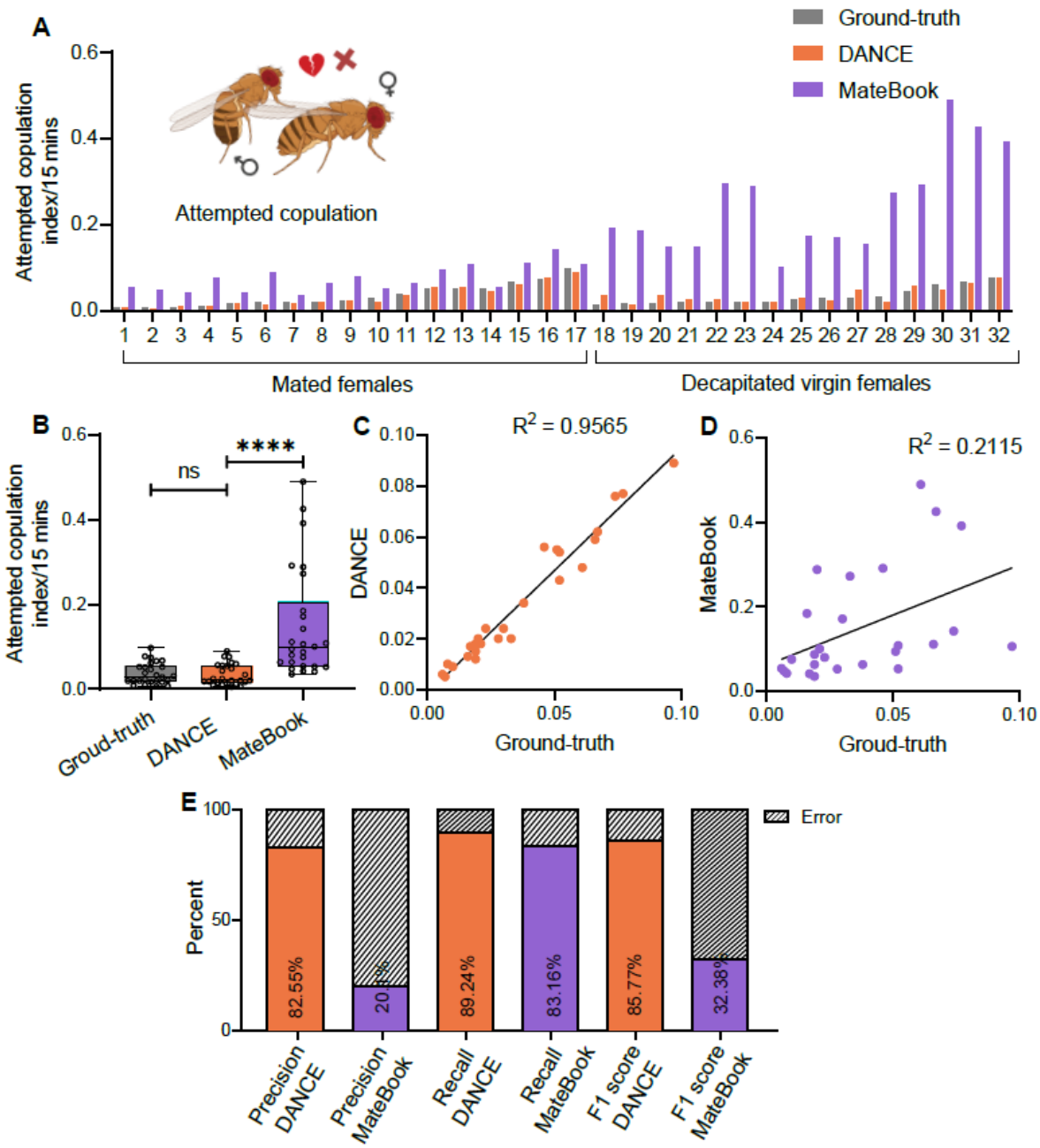


Figure 4. Validation of the DANCE attempted-copulation classifier. (A) Attempted copulation index of males from 15-minute-long videos scored by manual ground-truth (gray), the 'DANCE attempted copulation classifier' (orange) and the MateBook (purple) with both mated and decapitated females. (B) Comparison of ground-truth, DANCE attempted-copulation classifier, and MateBook scores (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=1512, n=32). (E) Precision, recall, and F1 score of the DANCE and MateBook attempted-copulation classifiers relative to the ground-truth scores.

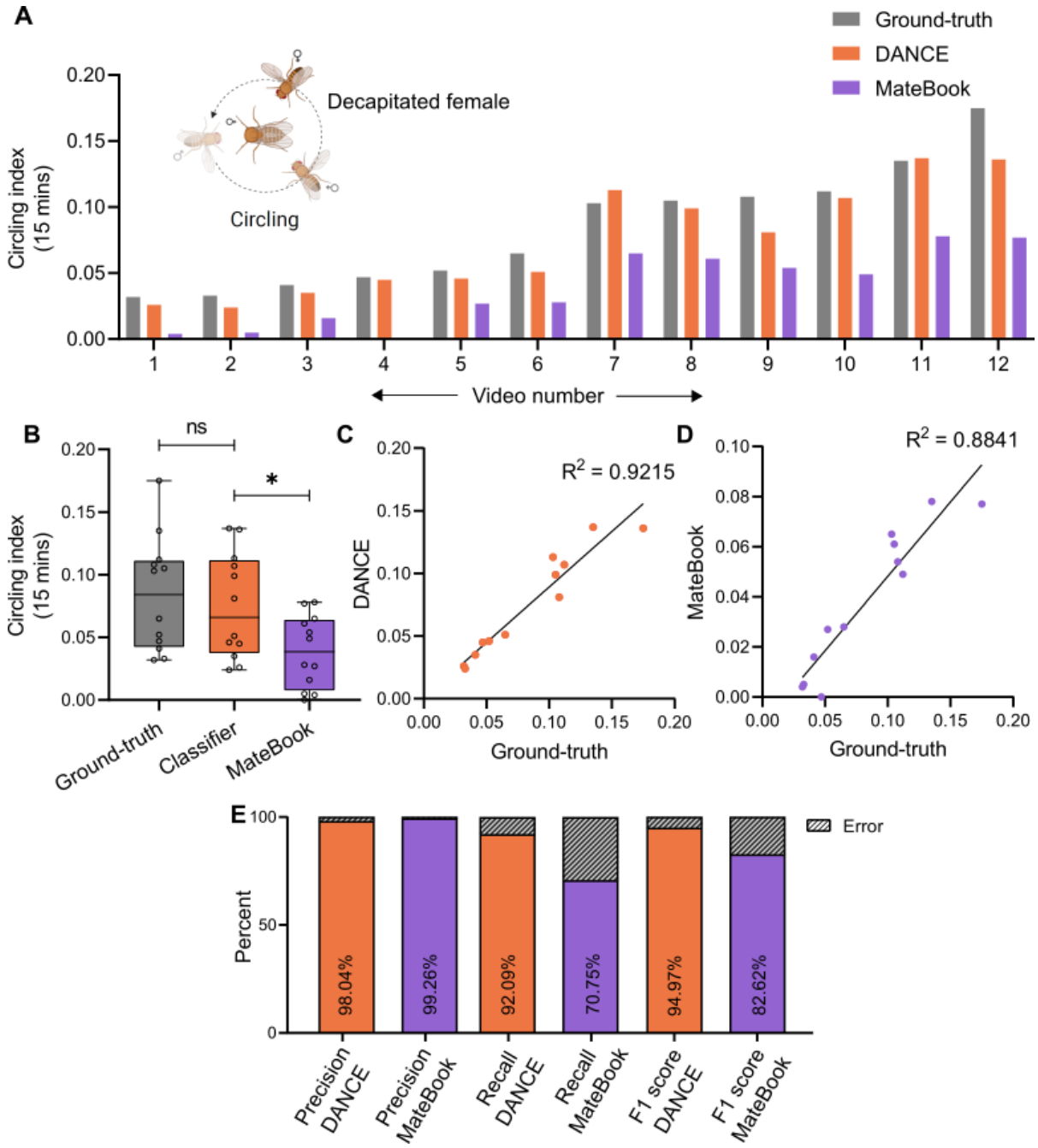


Figure 5. Evaluation of the DANCE circling classifier. (A) Circling index of males from 15 minute long videos from the ground-truth (gray), 'DANCE circling classifier' (orange) and MateBook (purple) with decapitated virgin females Friedman's ANOVA with Dunn's test, p=0.2049, ****p<0.0001, n=12. (B) Comparison of the ground-truth, DANCE, and MateBook circling classifiers (ordinary one-way ANOVA with Dunnett's test, ns, p=0.8014; *p=0.0157, n=12). (C) Regression analysis of the DANCE circling classifier *vs.* ground-truth (R^2=0.92, n=12). (D) Regression of MateBook *vs.* ground-truth (R^2=0.88, n=12). (E) Precision, recall, and F1 score of the DANCE and MateBook circling classifiers relative to the ground-truth score.

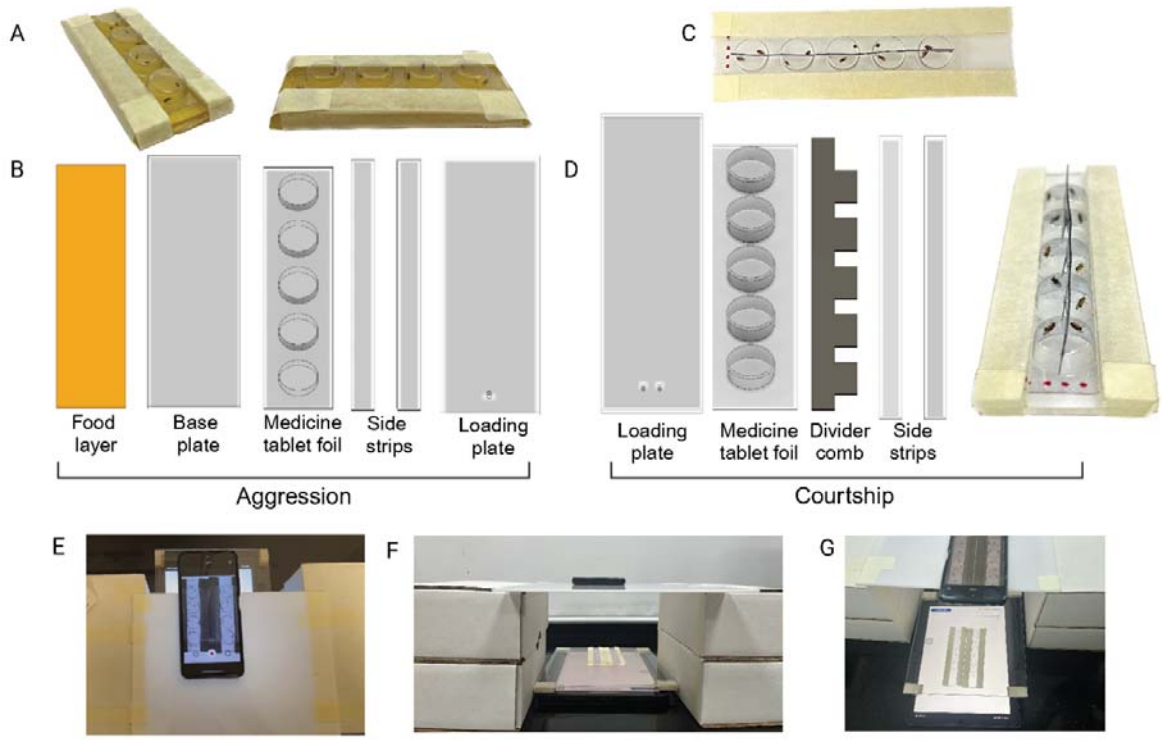
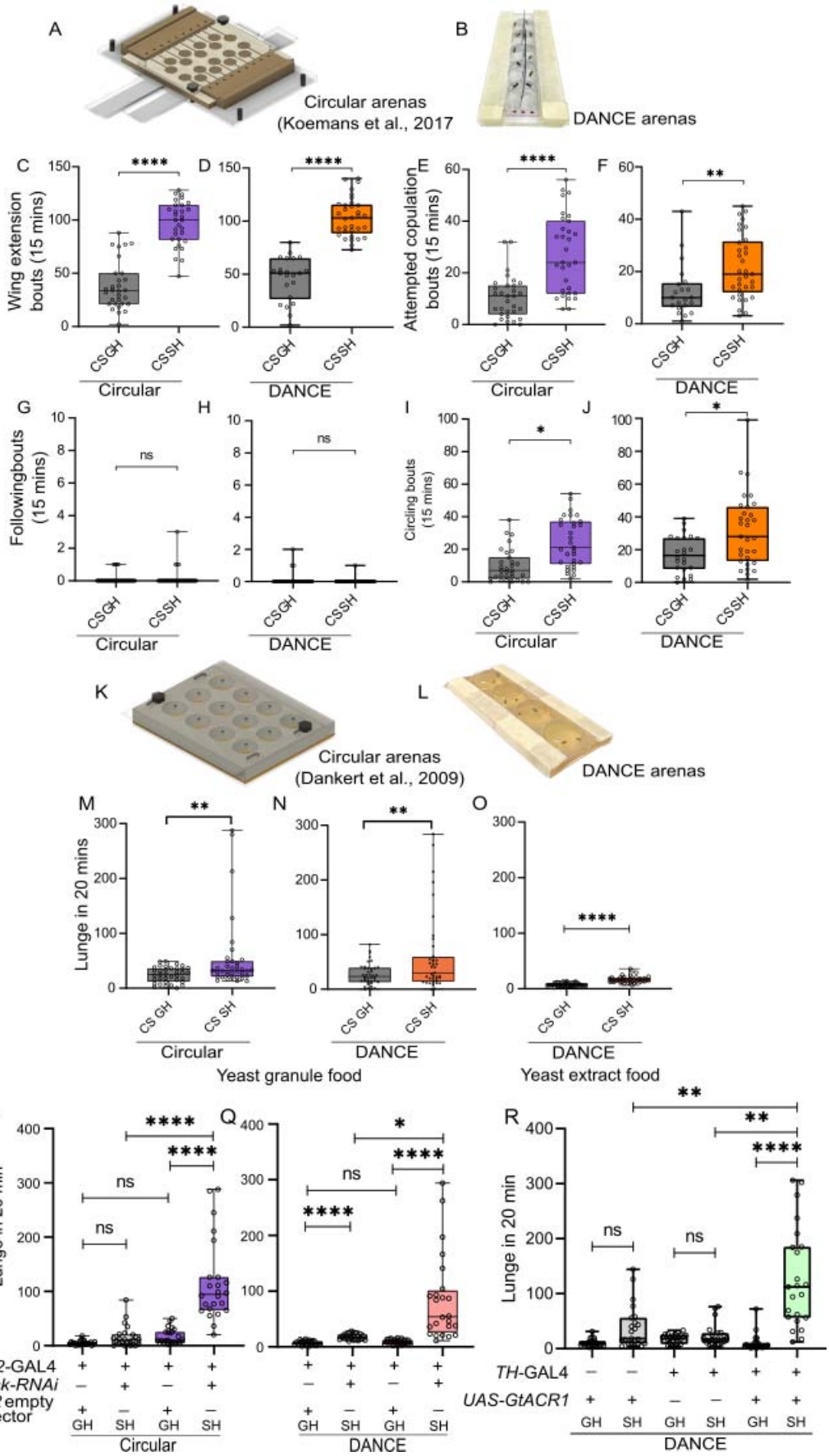


Figure 6. DANCE hardware and recording setup. (A) DANCE aggression setup (B) 3D-rendered components of the aggression setup (C) DANCE courtship setup (D) 3D-rendered components of the courtship setup, showing males and females separated by an X-ray film separator or 'divider comb'. (E-G) Top and side views of the DANCE setup with a smartphone camera for recording and an electronic tablet as the backlight.


A
Circular arenas
(Koemans et al., 2017
B
DANCE arenas
C
D
E
F
Wing extension bouts (15 mins)
Attempted copulation bouts (15 mins)
CSGH
CSSH
Circular
DANCE
G
H
I
J
Followingbouts (15 mins)
Circling bouts (15 mins)
ns
K
L
Circular arenas
(Dankert et al., 2009)
DANCE arenas
M
N
O
Lunge in 20 mins
CS GH
CS SH
Circular
DANCE
Yeast granule food
Yeast extract food
P
Q
R
Lunge in 20 min
Dilp2-GAL4
Dsk-RNAi
attp2 empty vector
TH-GAL4
UAS-GtACR1
GH
SH

752

Figure 7. Benchmarking DANCE hardware and application to neurogenetic tools. (A-B) Courtship behaviors recorded using a pre-existing circular setup and DANCE setup in group-housed (GH) and single-housed (SH) flies for (C-D) wing extension, (C) ***p<0.0010, n=23; (D) ****p<0.0001, GH, n=22 and SH, n=33. (E-F) Attempted copulation, (E) ***p<0.0002, n=23; (F) **p<0.0022, GH, n=21 and SH, n=33. (G-H) Following, (G) ns, p>0.0959, n=23; (H) ns, p<0.2537, GH, n=22 and SH, n=32. (I-J) Circling, (I) *p<0.012, n=23; (J) *p<0.0104, GH, n=24 and SH, n=31. (K-L) Aggressive lunges were recorded using a pre-existing circular setup and a DANCE setup. (M□N) Lunges of SH flies compared with those of 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 with *UAS-GtACR1* driven by the *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.

Supplementary information

Supplementary Figures and Legends

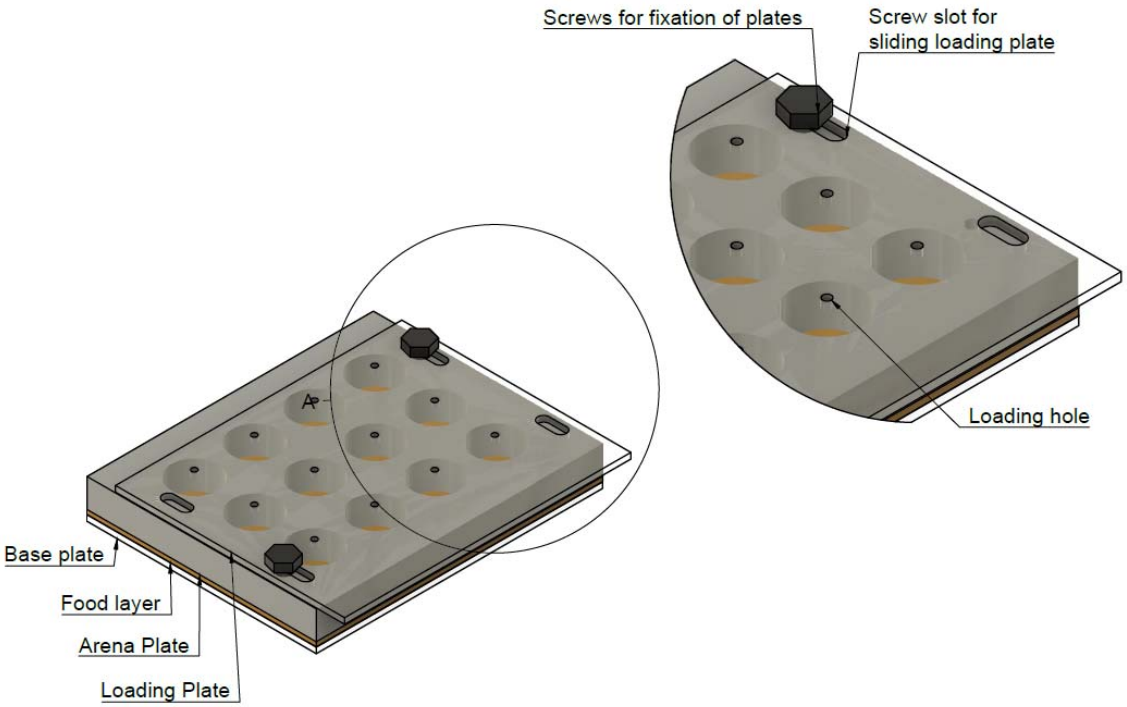


Figure 1— Figure Supplement 1. Aggression chamber described by Dankert et al. (2009). The setup consists of a bottom food plate, 12 wells (aggression arenas), and a top

plate with fly-loading holes and a screw slot that allows sliding of the loading plate.

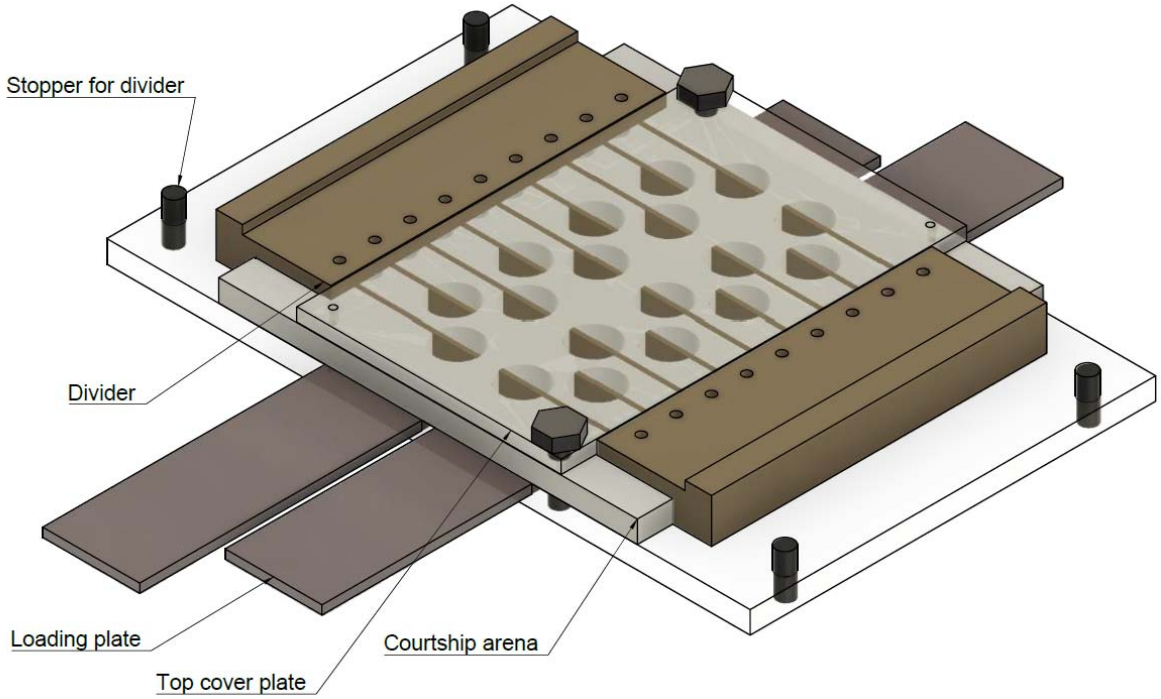


Figure 1— Figure Supplement 2. Courtship setup described by Koemans et al. (2017). The setup consists of 18 wells (courtship arenas), a top cover plate, a sliding loading plate, and a sliding divider assembly used to separate male and female flies.

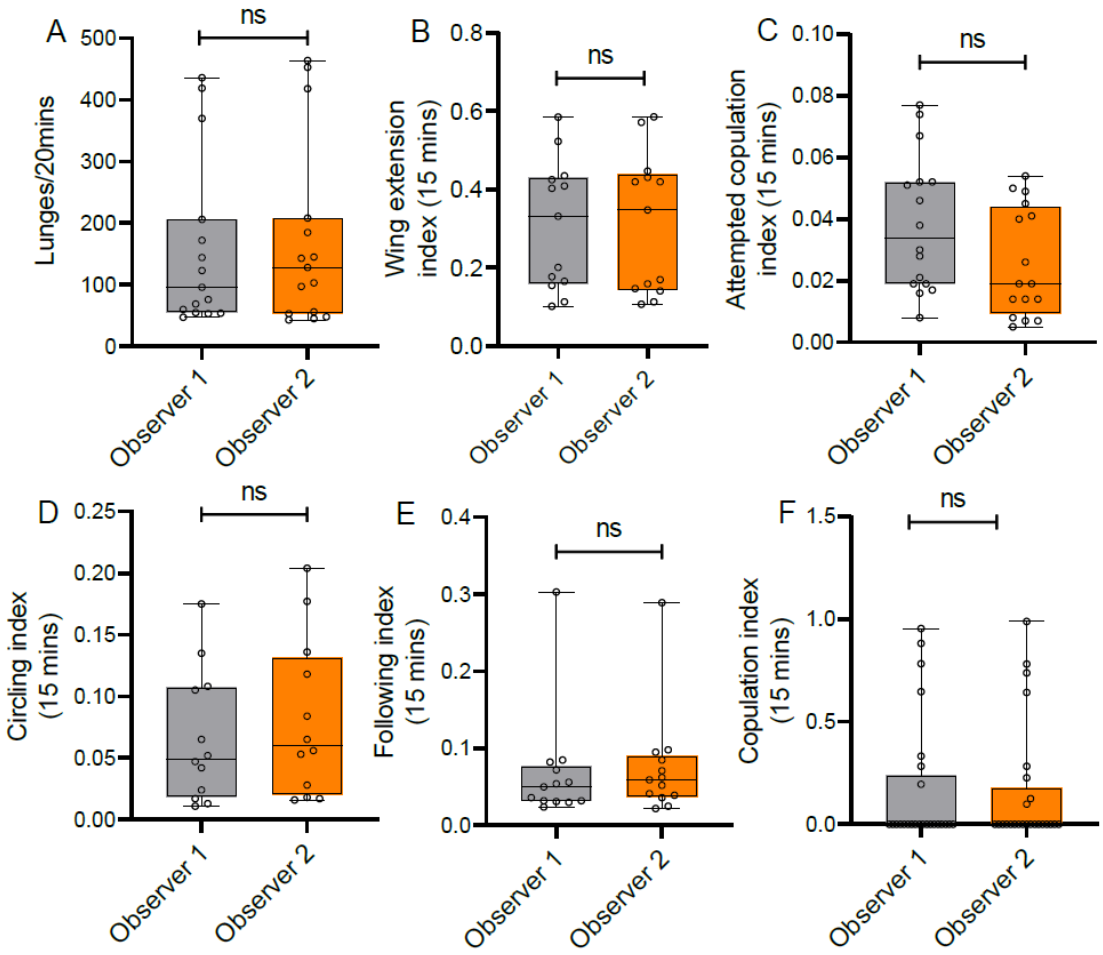


Figure 2— Figure Supplement 1. Comparison of annotations by two independent evaluators to assess observer bias during ground-truthing. (A) Aggressive lunges, p=0.8789, n=15. (B) Courtship wing extension, p=0.9999, n=13. (C) Attempted copulation, p=0.0571, n=16. (D) Circling, p=0.4343, n=12. (E) Following, p=0.4405, n=13. (F) Copulation, p=0.9221, n=25. (A-F) All comparisons were performed using the Mann□Whitney U test.

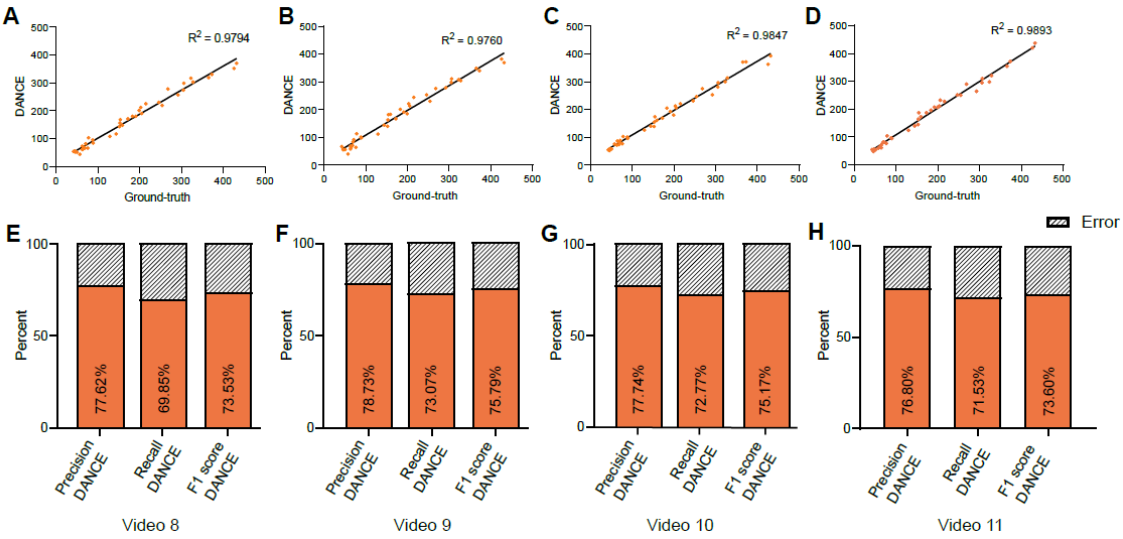


Figure 2— Figure Supplement 2. Evaluation of the DANCE lunge classifier predictions across training videos. (A–D) Scatter plots showing the correlation between manual ground-truth annotations and DANCE-predicted frame counts across four independent training videos (Videos 8–11). (A) $R^2 = 0.9794$, (B) $R^2 = 0.9760$, (C) $R^2 = 0.9847$, and (D) $R^2 = 0.9893$, demonstrating close agreement between automated classification and manual scoring. (E–H) Bar plots showing the precision, recall, and F1 score of the DANCE classifier for the corresponding videos**.** Video 9, which yielded the highest overall performance, was also used in Figure 2 for inter-method comparison with CADABRA and the Divider Assay to maintain consistency across analyses.

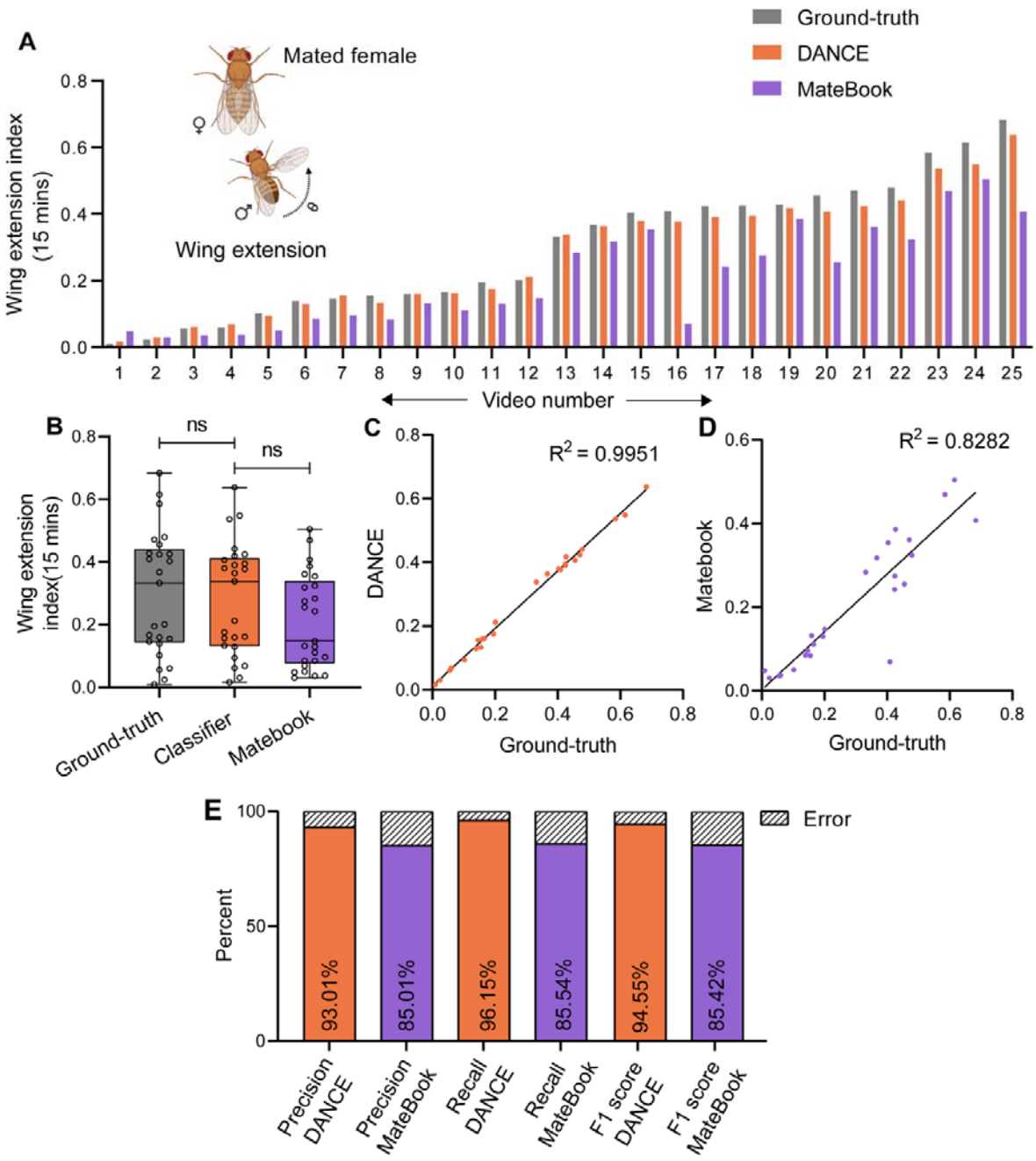


Figure 3— Figure Supplement 1. Evaluation of the DANCE wing extension classifier in the mated-female dataset.

(A) Wing extension of males from 15-minute videos scored by manual ground-truth (gray), DANCE classifier (orange), and MateBook (purple) with mated females. MateBook underscored wing extension across multiple videos (Friedman's ANOVA with Dunn's test; p=0.3582, p<0.0001; n=25). (B) Comparison of wing extension scores from ground-truth,

DANCE, and MateBook datasets (Kruskal–Wallis ANOVA with Dunn's test; $p>0.9999$, $p=0.1039$; $n=25$). (C) Regression of DANCE classifier scores vs. ground-truth ($R^2=0.9951$, $n=25$). (D) Regression of MateBook vs. ground-truth ($R^2=0.8282$, $n=25$). (E) Precision, recall, and F1 scores of the DANCE classifier and MateBook relative to ground-truth annotations.

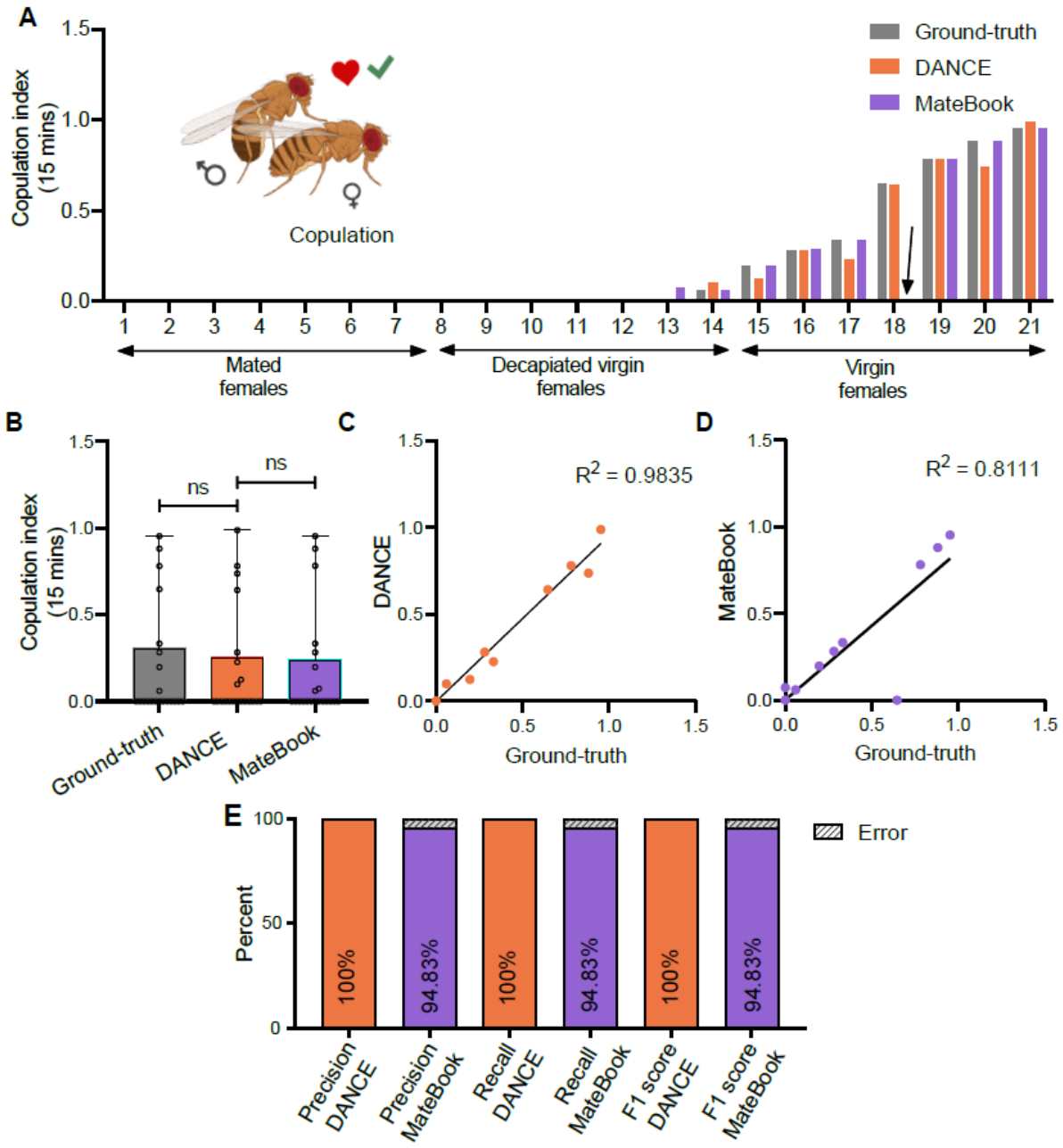


Figure 4— Figure Supplement 1. Evaluation of the DANCE copulation classifier in the mixed-female dataset. (A) Copulation scores from 15-min videos obtained by manual ground-truth (gray), the DANCE copulation classifier (orange), and MateBook (purple) (Friedman's ANOVA with Dunn's test, p>0.9999; p>0.9999, n=21). (B) Box plot comparison of the manual ground-truth, DANCE copulation classifier, and MateBook (Kruskal□Wallis ANOVA with Dunn's test, p>0.9999; p>0.9999, n=21). (C) Regression of

DANCE classifier scores vs. manual ground-truth (R^2=0.98, n=21).. (D) Regression of MateBook vs. manual ground-truth (R^2=0.81, n=21). (E) Precision, recall, and F1 score of the DANCE copulation classifier and MateBook relative to the ground-truth score.

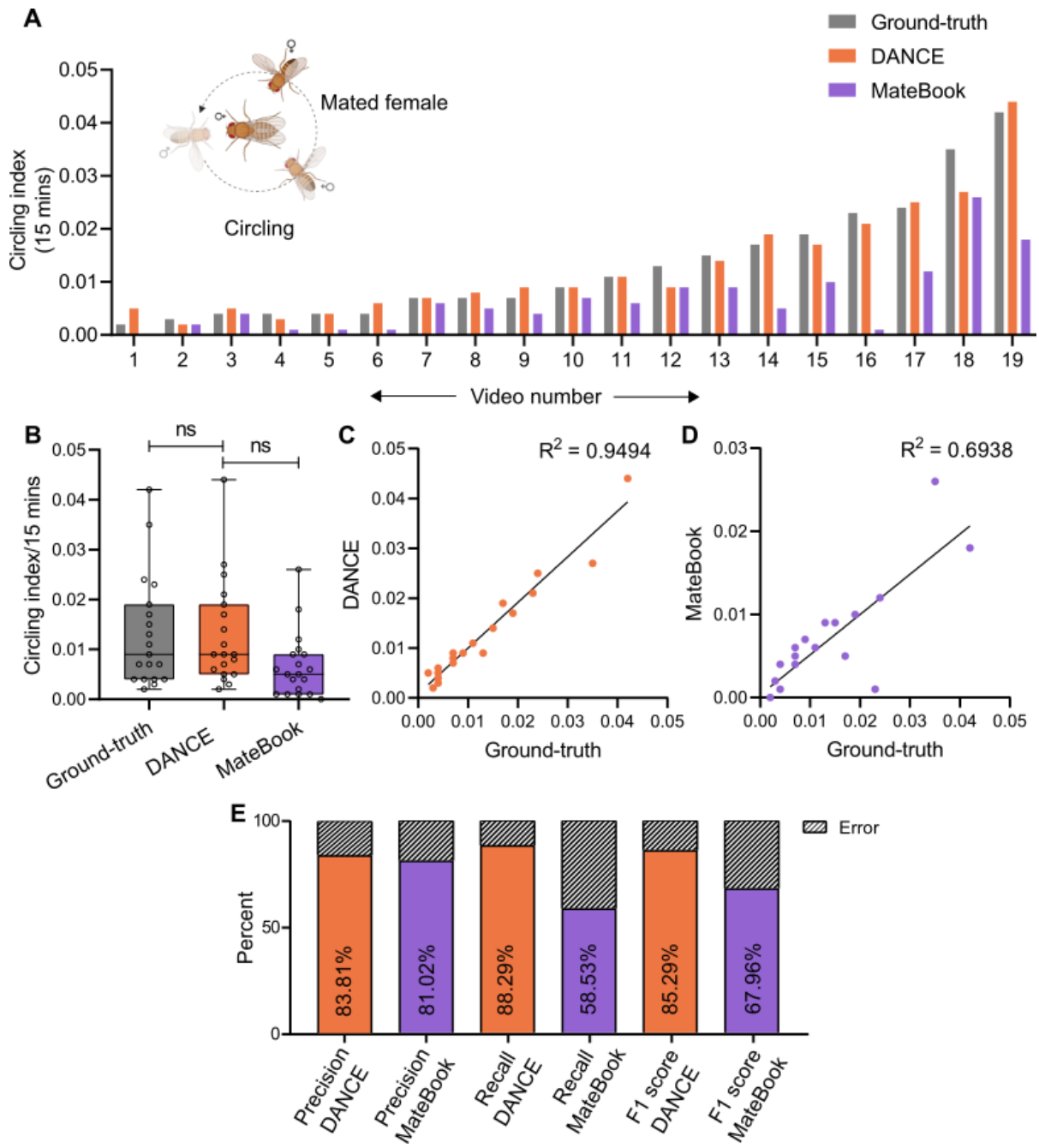


Figure 5— Figure Supplement 1. Evaluation of the DANCE circling classifier in the mated-female dataset. (A) Circling index of males from 15-minute videos scored by manual ground-truth (gray), the DANCE circling classifier (orange) and the MateBook (purple) with mated females (Friedman's ANOVA with Dunn's test, p>0.9999; p <0.0001, n=19). (B) Comparison of circling scores from ground-truth, DANCE, and MateBook datasets (Kruskal□Wallis ANOVA, p>0.9999; p=0.0822, n=19). (C) Regression of

DANCE classifier scores vs. ground-truth (R2 = 0.9494, n = 19). (D) Regression of MateBook vs. ground-truth (R^2=0.6938, n=19). (E) Precision, recall, and F1 scores of the DANCE and MateBook circling classifiers relative to the ground-truth scores.

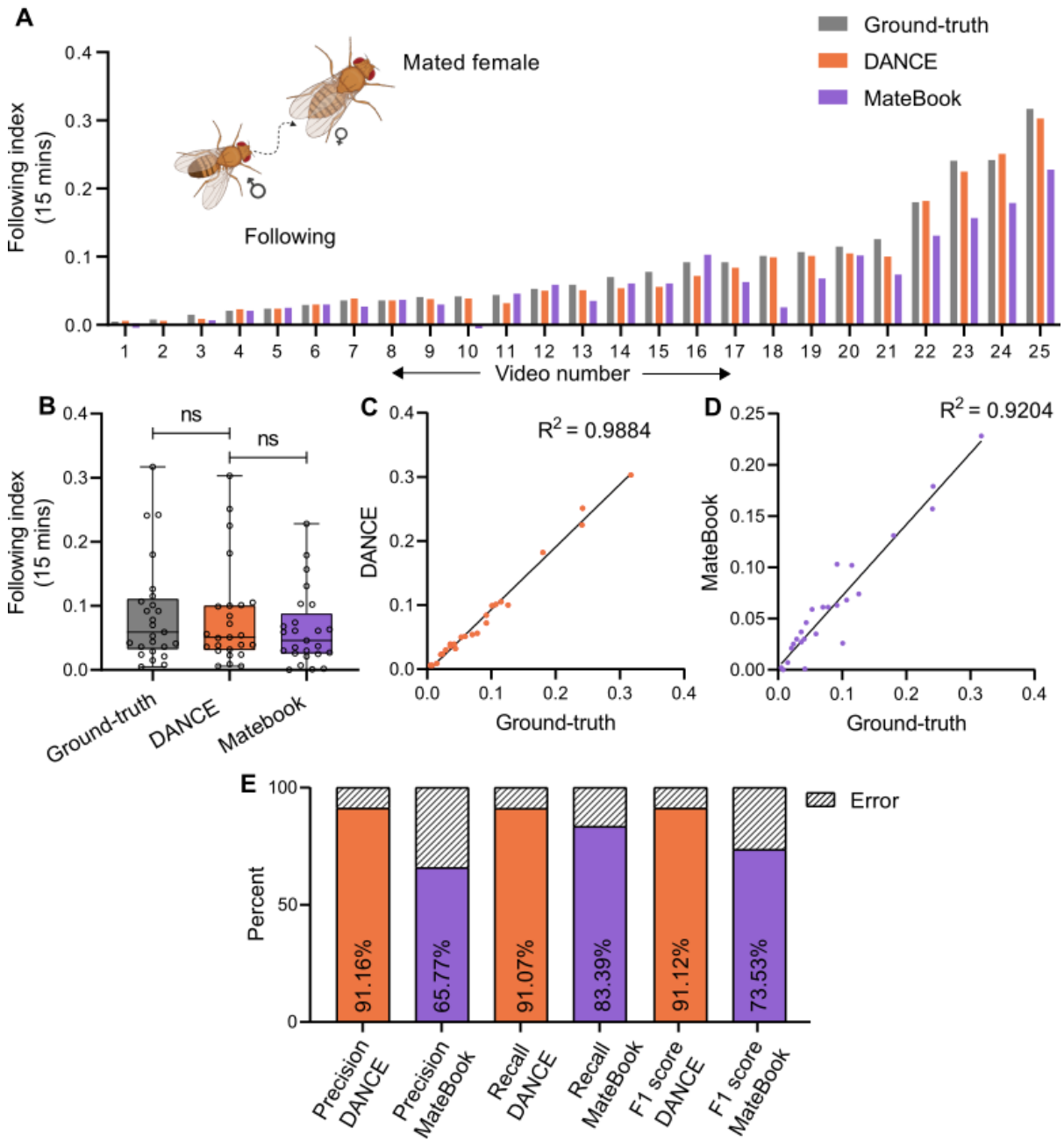
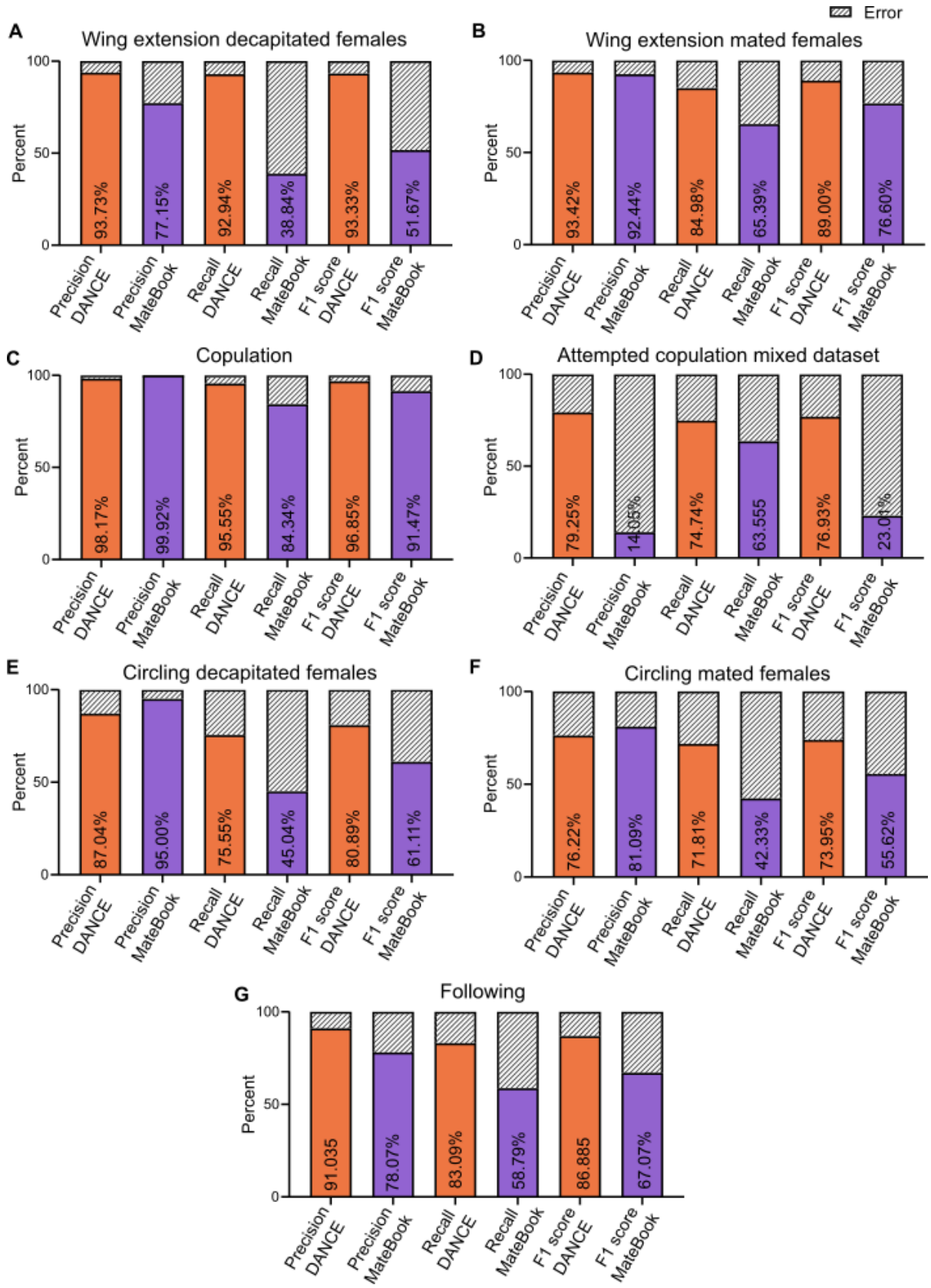


Figure 5— Figure Supplement 2. Evaluation of the DANCE following classifier in the mated-female dataset. (A) Following index of males from 15-minute videos scored by manual ground-truth (gray), DANCE following classifier (orange) and MateBook (purple) with mated females (Friedman's ANOVA with Dunn's test, p=0.1794; p=0.0029; n=25). (B) Box plot comparison of the following scores from the ground-truth, DANCE, and MateBook datasets (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) Regression of MateBook *vs.* ground-truth (R^2=0.9204, n=25). (E) Precision, recall, and F1 scores of DANCE and MateBook following classifiers relative to ground-truth annotations.


Error
A
Wing extension decapitated females
Percent
100
50
0
93.73%
77.15%
92.94%
38.84%
93.33%
51.67%
Precision DANCE
Precision MateBook
Recall DANCE
Recall MateBook
F1 score DANCE
F1 score MateBook
B
Wing extension mated females
93.42%
92.44%
84.98%
65.39%
89.00%
76.60%
C
Copulation
98.17%
99.92%
95.55%
84.34%
96.85%
91.47%
D
Attempted copulation mixed dataset
79.25%
14.05%
74.74%
63.555
76.93%
23.01%
E
Circling decapitated females
87.04%
95.00%
75.55%
45.04%
80.89%
61.11%
F
Circling mated females
76.22%
81.09%
71.81%
42.33%
73.95%
55.62%
G
Following
91.035
78.07%
83.09%
58.79%
86.885
67.07%

848

Figure 5— Figure Supplement 3. Frame-level analysis of duration-based courtship classifiers. (A-G) Bar plots showing precision, recall, and F1 scores for DANCE (orange) and MateBook (purple) across different behaviors: (A) Wing extension toward decapitated females, (B) Wing extension toward mated females, (C) Copulation, (D) Attempted copulation (mixed dataset), (E) Circling toward decapitated females, (F) Circling toward mated females, and (G) Following. Percent error rates are shown above each bar.

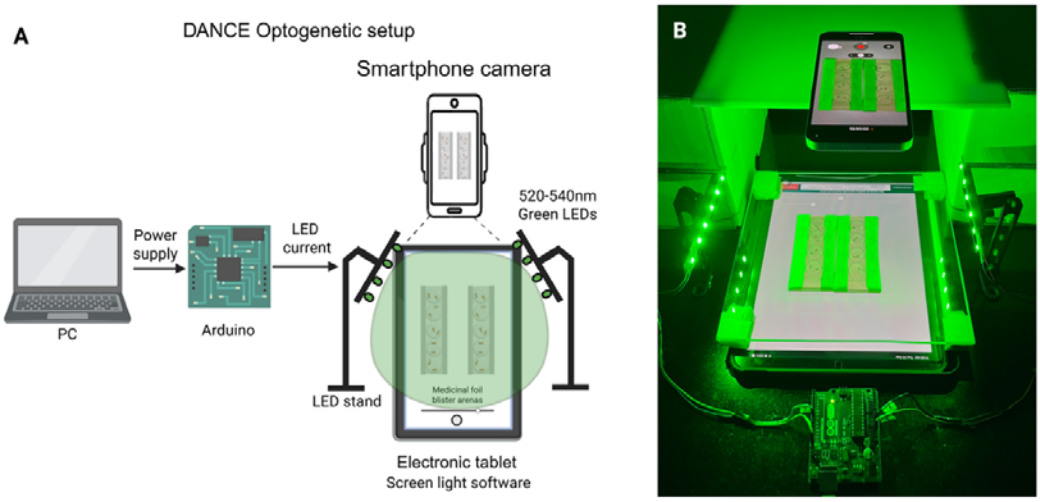


Figure 7—Figure Supplement 1. DANCE optogenetic recording setup for behavioral experiments. (A) Schematic of the setup showing green LEDs (520–540 nm) controlled by an Arduino and powered via a PC, with illumination directed onto blister-pack arenas placed on an electronic tablet. The tablet provides an adjustable backlight via screen-light software, and a smartphone camera records behavior. (B) Photograph of the complete setup, showing the smartphone, tablet, LED stands, Arduino controller, and arenas under green LED illumination.

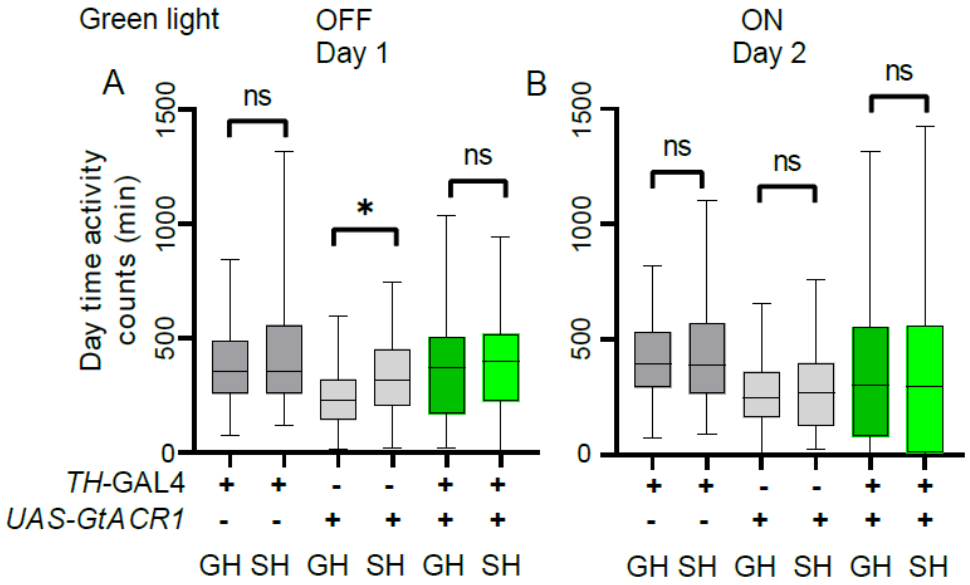


Figure 7— Figure Supplement 2. Effect of optogenetic silencing of dopaminergic neurons on daytime activity. (A–B) Transient silencing of dopaminergic neurons using *UAS-GtACR1* did not affect daytime activity between SH and GH flies. (A) On day 1 (no green-light silencing), one-way ANOVA with Tukey's multiple comparisons revealed no differences for *TH-GAL4* (ns, p=0.7383; GH, n=43; SH, n=43), *UAS-GtACR1* (ns, p=0.4812; GH, n=51; SH, n=42), and *TH-GAL4>UAS-GtACR1* (ns, p=0.9942; GH, n=54; SH, n=51). (B) On day 2 (with green-light silencing), no differences were observed for *TH-GAL4* (ns, p=0.9976; GH, n=43; SH, n=43), *UAS-GtACR1* (ns, p=0.9779; GH, n=51; SH, n=42), and *TH-GAL4>UAS-GtACR1* (ns, p=0.9974; GH, n=54; SH, n=51). Two-way ANOVA across days revealed no significant interaction or main effects for *TH-GAL4* (interaction: ns, p=0.5504; silencing: ns, p=0.5172; housing: ns, p=0.1602), *UAS-GtACR1* (interaction: ns, p=0.4533; silencing: ns, p=0.3602; housing: p=0.255; GH, n=51; SH, n=42), or *TH-GAL4>UAS-GtACR1* (interaction: ns, p=0.9977; silencing: ns, p=0.2454; housing: ns, p=0.5868). Within-day comparisons were performed using one-way ANOVA

with Tukey's multiple comparisons; across-day comparisons were performed using two-way ANOVA.

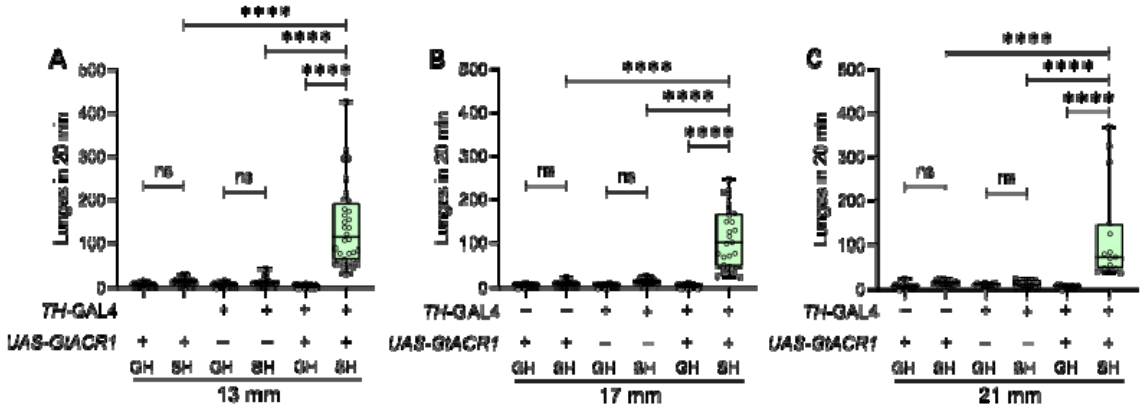
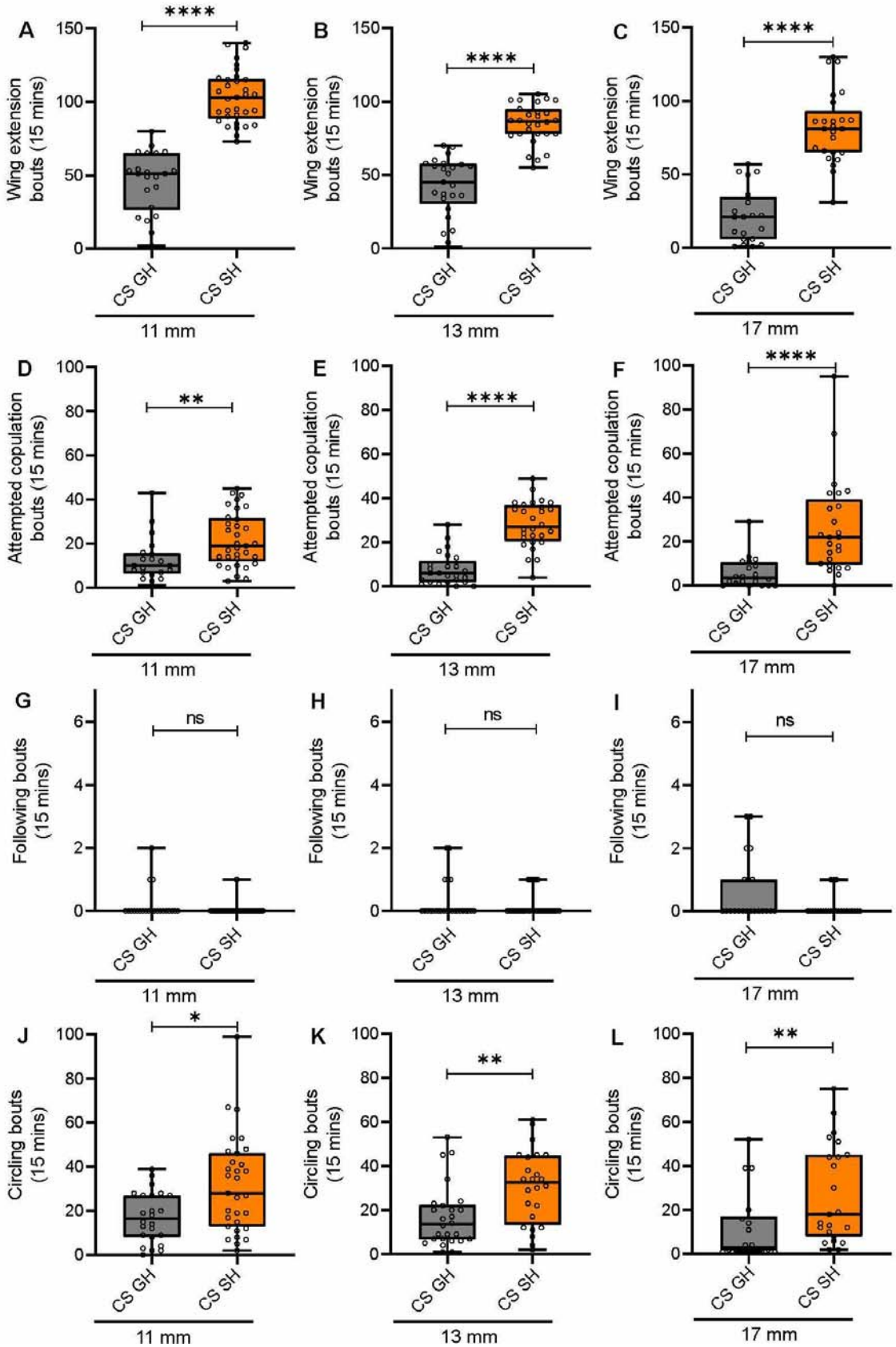


Figure 7—Figure Supplement 3. Effect of optogenetic silencing of dopaminergic neurons on male aggression across arena sizes. (A–C) Number of lunges performed in 20 minutes by *TH-GAL4>UAS-GtACR1* males tested in arenas of different diameters: (A) 13 mm (n=29–38), (B) 17 mm (n=27–32), and (C) 21 mm (n=16). Lunges of flies with *GtACR1*-mediated silencing were significantly greater than those of GH and control flies across all arena sizes. Statistical comparisons are indicated: ****p<0.0001; ns, not significant; ordinary one-way ANOVA followed by Tukey's multiple comparisons test.


A
B
C
D
E
F
G
H
I
J
K
L
Wing extension bouts (15 mins)
Attempted copulation bouts (15 mins)
Following bouts (15 mins)
Circling bouts (15 mins)
****
**
ns
*
CS GH
CS SH
11 mm
13 mm
17 mm

892

Figure 7—Figure Supplement 4. Quantification of wild-type *Drosophila* courtship behavior in DANCE chambers of varying diameters. (A–C) Wing extension bouts (A: n=22, 33; B: n=25, 28; C: n=20, 25). (D–F) Attempted copulation bouts (D: n=21, 33; E: n=25, 28; F: n=20, 25). (G–I) Following bouts (G: n=22, 32; H: n=25, 28; I: n=21, 24). (J–L) Circling bouts (J: n=24, 31; K: n=26, 24; L: n=18, 23). Data were obtained from Canton-S (CS) males paired with either group-housed (GH) or single-housed (SH) females over 15 minutes in chambers of 11 mm, 13 mm, or 17 mm diameter. Each bar represents the median. Statistical differences are indicated as *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001; ns, not significant (Mann–Whitney U test).

References

Agrawal P, Kao D, Chung P, Looger LL. 2020. The neuropeptide Drosulfakinin regulates social isolation-induced aggression in Drosophila. *J Exp Biol* **223**:jeb207407.

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.

Anderson DJ, Perona P. 2014. Toward a science of computational ethology. *Neuron* **84**:18–31.

Asahina K. 2017. Neuromodulation and Strategic Action Choice in Drosophila Aggression. *Annu Rev Neurosci* **40**:51–75.

Asahina K, Watanabe K, Duistermars BJ, Hoopfer E, González CR, Eyjólfsdóttir EA, Perona P, Anderson DJ. 2014. Tachykinin-expressing neurons control male-specific aggressive arousal in drosophila. *Cell* **156**:221–235.

Bastock M, Manning A. 1955. The Courtship of Drosophila Melanogaster. *Behaviour* **8**:85–110.

Bennet-Clark HC, Ewing AW. 1969. Pulse interval as a critical parameter in the courtship song of Drosophila melanogaster. *Anim Behav* **17**:755–759.

Bentzur A, Ben-Shaanan S, Benichou JIC, Costi E, Levi M, Ilany A, Shohat-Ophir G. 2021. Early Life Experience Shapes Male Behavior and Social Networks in Drosophila. *Curr Biol* **31**:486–501.e3.

Benzer S. 1967. BEHAVIORAL MUTANTS OF Drosophila ISOLATED BY COUNTERCURRENT DISTRIBUTION. *Proc Natl Acad Sci U S A* **58**:1112–1119.

Branson K, Robie AA, Bender J, Perona P, Dickinson MH. 2009. High-throughput ethomics in large groups of Drosophila. *Nat Methods* **6**:451–457.

Brown RG. 1965. Courtship behaviour in the Drosophila obscura group. II. Comparative studies. *Behaviour* **25**:281–323.

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.

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.e16.

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.

Cook R, Cook A. 1975. The attractiveness to males of female Drosophila melanogaster: effects of mating, age and diet. *Anim Behav* **23**:521–526.

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.

Datta SR, Anderson DJ, Branson K, Perona P, Leifer A. 2019. Computational Neuroethology: A Call to Action. *Neuron* **104**:11–24.

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.

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.

Dickson BJ. 2008. Wired for sex: the neurobiology of Drosophila mating decisions. *Science* **322**:904–909.

Dierick HA. 2007. A method for quantifying aggression in male drosophila melanogaster. *Nat Protoc* **2**:2712–2718.

Dow MA, Schilcher F von. 1975. Aggression and mating success in Drosophila melanogaster. *Nature* **254**:511–512.
Duistermars BJ, Pfeiffer BD, Hoopfer ED, Anderson DJ. 2018. A Brain Module for Scalable Control of Complex, Multi-motor Threat Displays. *Neuron* **100**:1474–1490.e4.
Eyjolfsdottir E, Branson S, Burgos-Artizzu XP, Hoopfer ED, Schor J, Anderson DJ, Perona P. 2014. Detecting social actions of fruit flies. *Computer Vision – ECCV* **8690**:772–787.
Fowler GL. 1973. Some Aspects of the Reproductive Biology of Drosophila: Sperm Transfer, Sperm Storage, and Sperm Utilization In: Caspari EW, editor. Advances in Genetics. Academic Press. pp. 293–360.
Gill KS. 1963. A mutation causing abnormal courtship and mating behavior in males of Drosophila melanogaster. *Am Zool* **3**:507.
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, Fernandez J, Martin FA. 2023. A simplified courtship conditioning protocol to test learning and memory in Drosophila. *STAR Protocols* **4**. doi:10.1016/j.xpro.2022.101572
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.
Govorunova EG, Sineshchekov OA, Janz R, Liu X, Spudich JL. 2015. NEUROSCIENCE. Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics. *Science* **349**:647–650.
Greenspan RJ, Ferveur JF. 2000. Courtship in Drosophila. *Annu Rev Genet* **34**:205–232.
Hall JC. 2002. Courtship lite: a personal history of reproductive behavioral neurogenetics in Drosophila. *J Neurogenet* **16**:135–163.
Hall JC. 1978. Courtship among males due to a male-sterile mutation in Drosophila melanogaster. *Behav Genet* **8**:125–141.
Hindmarsh Sten T, Li R, Hollunder F, Eleazer S, Ruta V. 2025. Male-male interactions shape mate selection in Drosophila. *Cell* **188**:1486–1503.e25.
Hoffmann AA. 1987a. A laboratory study of male territoriality in the sibling species Drosophila melanogaster and D. simulans. *Anim Behav* **35**:807–818.
Hoffmann AA. 1987b. Territorial Encounters Between Drosophila Males of Different Sizes. *Anim Behav* **35**:1899–1901.
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.
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.
Ishii K, Cortese M, Leng X, Shokhirev MN, Asahina K. 2022. A neurogenetic mechanism of experience-dependent suppression of aggression. *Science Advances* **8**. doi:10.1126/sciadv.abg3203
Jacobs ME. 1960. Influence of Light on Mating of Drosophila Melanogaster. *Ecology* **41**:182–188.
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**. doi:10.1038/s41467-021-23041-y
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.e5.
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.
Kessler S. 1962. Courtship rituals and reproductive isolation between the races or incipient species of Drosophila paulistorum. *Am Nat* **96**:117–121.
Kim Y-K, Ehrman L. 1998. Developmental Isolation and Subsequent Adult Behavior of Drosophila paulistorum.IV. Courtship. *Behav Genet* **28**:57–65.
Kim YKY-K, Saver M, Simon J, Kent CFCF, Shao L, Eddison M, Agrawal P, Texada M, Truman JWJW, 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.
Koemans TS, Oppitz C, Donders RAT, Van Bokhoven H, Schenck A, Keleman K, Kramer JM. 2017. Drosophila courtship conditioning as a measure of learning and memory. *J Vis Exp* **2017**:1–11.
Kravitz EA, Fernandez M de la P. 2015. Aggression in Drosophila. *Behav Neurosci* **129**:549–563.
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**:1–15.
Lim RS, Eyjólfsdóttir E, Shin E, Perona P, Anderson DJ. 2014. How food controls aggression in Drosophila. *PLoS One* **9**.
Lopes G, Bonacchi N, Frazão J, Neto JP, Atallah BV, Soares S, Moreira L, Matias S, Itskov PM, Correia PA, Medina RE, Calcaterra L, Dreosti E, Paton JJ, Kampff AR. 2015. Bonsai: an event-based framework for processing and controlling data streams. *Front Neuroinform* **9**:7.
MacBean IT, Parsons PA. 1967. Directional selection for duration of copulation in Drosophila melanogaster. *Genetics* **56**:233–239.
Markow TA. 1987. Behavioral and sensory basis of courtship success in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **84**:6200–6204.
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.
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.
Nilsen SP, Chan YB, Huber R, Kravitz EA. 2004. Gender-selective patterns of aggressive behavior in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **101**:12342–12347.
O'Dell KMC. 2003. The voyeurs' guide to Drosophila melanogaster courtship. *Behav Processes* **64**:211–223.
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.
Pan Y, Baker BS. 2014. Genetic Identification and Separation of Innate and Experience-Dependent Courtship Behaviors in Drosophila. *Cell* **156**:236–248.
Pavlou HJ, Goodwin SF. 2013. Courtship behavior in Drosophila melanogaster: towards a "courtship connectome." *Curr Opin Neurobiol* **23**:76–83.
Reza MA, 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 (Austin)* **7**:105–111.
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.e18.

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. *Cell* **170**:393–406.e28.

Robie AA, Seagraves KM, Egnor SER, Branson K. 2017b. Machine vision methods for analyzing social interactions. *J Exp Biol* **220**:25–34.

Seeds AM, Ravbar P, Chung P, Hampel S, Midgley FM Jr, Mensh BD, Simpson JH. 2014. A suppression hierarchy among competing motor programs drives sequential grooming in Drosophila. doi:10.7554/eLife.02951

Shorey HH. 1962. Nature of the Sound Produced by Drosophila melanogaster during Courtship. *Science* **137**:677–678.

Simon JC, Heberlein U. 2020. Social hierarchy is established and maintained with distinct acts of aggression in male. *J Exp Biol* **223**. doi:10.1242/jeb.232439

Sokolowski MB. 2001. Drosophila: Genetics meets behaviour. *Nature Reviews Genetics* **2**:879–890.

Spieth HT. 1974. Courtship behavior in Drosophila. *Annu Rev Entomol* **19**:385–405.

Spieth HT. 1968. Evolutionary Implications of Sexual Behavior in Drosophila In: Dobzhansky T, Hecht MK, Steere WC, editors. Evolutionary Biology: Volume 2. Boston, MA: Springer US. pp. 157–193.

Spieth HT. 1966. Drosophilid mating behavior: the behaviour of decapitated females. *Anim Behav* **14**:226–235.

Spieth HT. 1952. Mating Behavior Within the Genus Drosophila (Diptera).

Sturtevant AH. 1915. Experiments on sex recognition and the problem of sexual selection in Drosoophilia. *J Exp Psychol Anim Behav Process* **5**:351–366.

Swain B, von Philipsborn AC. 2021. Sound production in Drosophila melanogaster: Behaviour and neurobiologyAdvances in Insect Physiology. pp. 141–187.

Tao L, Ayambem D, Barranca VJ, Bhandawat V. 2024. Neurons Underlying Aggression-Like Actions That Are Shared by Both Males and Females in. *J Neurosci* **44**. doi:10.1523/JNEUROSCI.0142-24.2024

Tompkins L, Hall JC, Hall LM. 1980. Courtship-stimulating volatile compounds from normal and mutant Drosophila. *J Insect Physiol* **26**:689–697.

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.

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.e7.

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.

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.

Zhang N, Guo L, Simpson JH. 2020. Spatial Comparisons of Mechanosensory Information Govern the Grooming Sequence in Drosophila. *Curr Biol* **30**:3697–3698.