PNAS
www.pnas.org

Supplementary Information for

The Evolution of Social Organization: Behavioural and Environmental Contributions to Drosophilid Social Networks

Jacob A. Jezovit,[1] Rebecca Rooke,[1] Jonathan Schneider,[1] Joel D. Levine[1]

[1] Department of Biology, University of Toronto at Mississauga, Mississauga, Ontario, Canada L5L 1C6
* Joel D. Levine, 3359 Mississauga Rd., University of Toronto at Mississauga, Mississauga, Ontario, Canada L5L 1C6, (905) 828-5364

Email: joel.levine@utoronto.ca

This PDF file includes:

Supplementary text
Figures S1 to S12
Tables S1 to S4
SI References

www.pnas.org/cgi/doi/10.1073/pnas.1920642117

Fly stocks

The majority of the species used for the SIN assay were purchased from the *Drosophila* Species Stock Centre (https://blogs.cornell.edu/drosophila/orders/) as shown in Table S1. The Canton-S strain was used to represent *D. melanogaster*. The *D. yakuba* stock was a gift from N. Gompel. The source of the *D. mojavensis* and *D. sechellia* stocks are unknown, but these stocks were confirmed to be the correct species through documented phenotypic markers (1).

Video acquisition

SIN data were acquired as outlined previously (2). To summarize, all fly stocks were maintained in bottles with 40 mL of media containing cornmeal, wheat germ, soy flour, molasses, sucrose, glucose, yeast, agar, propionic acid and Tegosept. These bottles were stored in a Percival incubator (model I-36LL) set to 25°C with a 12 h L:D cycle. For all species, virgin male and female flies were collected through light CO_2 anaesthesia. Male and female flies were housed separately in groups of 12-16 individuals within vials containing 8 mL of media. All flies were reared for 3 days in an incubator controlled under the same temperature and photoperiod described above. Afterwards, 12 flies from each vial were gently aspirated into circular arenas (60 mm diameter, 2 mm depth) and filmed for 30 minutes using Firefly MV cameras (Point Grey). All videos were filmed at the same time of day (9.5-10.5 h after initiation of light phase during photoperiod) to control for the clock-controlled locomotor activity of flies. All filming was completed inside a Biochamber with controlled temperature (25°C) and humidity (65% RH). After filming, all flies were anaesthetized with CO_2 and discarded. The position, orientation and identity of all flies in each video were tracked through machine vision software (Ctrax; versions 0.4.2 & 0.5.19b). Errors in tracking were manually corrected through the fixerrors (version 0.2.3) MATLAB package to correct for: swapping of fly identities, errors in the orientation of the head to abdomen region of the fly tracks and drastic fluctuations in the large major axis of the fly tracks. Data collection was performed separately for groups of male flies and for groups of female flies and we will use the terms “male dataset” and “female dataset”, respectively. In total, we collected, tracked and fixed the errors in 1000 video trials, totaling approximately forty million frames of video, across the male and female datasets of the 20 drosophilid species. Social interaction network (SIN) data was acquired from 19 *Drosophila* species and one outgroup species, *Chymomyza procnemis*. We gathered data for 20 species since a power analysis reported this as a reliable minimum sample size for phylogenetic comparative methods (3, 4). Approximately 20-25 videos were acquired for each species in both the male and female datasets (see Figure S1 for precise sample sizes of each species).

Pilot data

An additional male and female dataset containing 399 videos across 7 species was collected and analyzed. This collection of videos was acquired between July 2014 and August 2015. The following species are represented in this pilot data: *D. melanogaster* (Canton-S), *D. simulans*, *D. sechellia*, *D. yakuba*, *D. pseudoobscura*, *D. mojavensis*, and *D. paramelanica*. With this pilot data, we were able to test the robustness of SIN measures between 5 species common to both independent data collections. The stocks for these species were the same as those used in the 20 species-wide datasets. The relative species differences, for each dataset, were statistically tested through a Kruskal-Wallis one-way ANOVA in MATLAB. The replication of these datasets was visualized through boxplots that display the distribution of the SIN scores for each species (Figure S5-S6).

Estimation of social interaction networks and other behavioural measures

We considered that different drosophilid species may interact differently. We considered how the inter-individual distance, angle and time, collectively referred to as the social spacing parameters

may vary across species. To account for this, we utilized an algorithm that estimates these parameters based on tracked videos (see below). Each species' distance, angle and time parameters for both the male and female datasets are listed in Table S2. These species-specific and sex-specific social spacing parameters allowed us to control for variation in social interactions when generating SINs. Additionally, the social spacing parameters can also be used as a measure to classify the ways that different species interact.

All SINs were generated as previously described (2). To summarize, all SINs are iterative networks that comprise 33 unique interactions, which represents a network density of 25% (25% of 132 unique possible interactions in a group of 12 individuals). The number of iterative networks in a single 30-minute video may vary and is proportional to the number of unique interactions in a video. Most species formed at least one SIN iteration in over 80% of their respective video trials (Figure S1). We assess four SIN measures that we view as social organization phenotypes: i) assortativity, defined as the probability of an individual interacting with another individual with a similar degree (degree is defined as the number of incoming and outgoing connections to a single node); ii) clustering coefficient, defined as a measure of how interconnected neighbours are to one another; iii) betweeness centrality, defined as the number of shortest paths that traverse an individual, indicating the relative importance of an individual for maintaining the cohesion of the network; and iv) global efficiency, defined as a measure of redundant pathways, indicating the efficiency of information flow throughout the network (2, 6). Here we speak about these network measures as they apply to a group on average. The four SIN measures are expressed as z-scores, which normalizes the networks to control for degree distribution. To do this, we generated 10000 random networks and calculated the z-score as follows:

$$\frac{measurement_{observed} - mean(measurement_{random})}{standard\ deviation\ (measurement_{random})}$$

For each SIN measure, we present the mean z-score, averaged from the distribution of videos for each species in the male and female dataset (Figure 3). Like the SIN measures, the mean of each behavioural element is calculated from the distribution of videos for each species in the male and female dataset (Figure 2).

In addition to the SIN measures, the following behavioural elements were measured: i) movement, defined as the mean locomotor activity (millimeters/second) of the flies in a single video; ii) interaction duration, defined as the mean duration of the social interactions (based on the distance, angle and time parameters) in a single video; iii) reciprocation, defined as the proportion of social interactions reciprocated in a single video; and iv) number of interactions, defined as the total number of social interactions in a single video

Automated estimation of social spacing interaction criteria

All interaction criteria used to generate SINs in the 20 species-wide data (referred to as experimental data below) are shown in Table S2. The distance, angle and time interaction criteria were estimated for each species using an algorithm published by Schneider and Levine (5). Stated briefly, the algorithm analyzes the spatial positions of every fly in all tracked videos. The algorithm eliminates background noise from flies by analyzing spatial positions of "virtual trials" which consist of fly tracks randomly sampled from separate videos. With that background subtraction, the algorithm identifies distance, angle and time criteria that are over-represented in videos of flies socially interacting compared to the non-social virtual trials. Initially, the algorithm provided interaction criteria that contradicted personal observations of each species' spatial patterning. The following adjustments were made to improve the performance of the algorithm:

1. Previously the algorithm computed an “inter-fly distance” defined as the interval where the distribution for distance measures was positive when subtracted from the virtual trials. This interval was the starting point to locate interaction hotspots on the social space heat map (see S1-S25 for a visualization of these heat maps). Here, we did not compute the inter-fly distance and established the initial interaction hotspots by targeting the 95th percentile instead of the third quartile. As reported previously, angle and distance bins were increased until the mean of the enclosed region on the heat maps began to decrease.
2. Previously the algorithm computed the time criteria by recording the time elapsed when flies fulfilled the calculated distance and angle criteria. Background noise would be removed by subtracting the distribution of time values from the virtual trial time distribution. The first positive time bin, representing the minimal time value over-represented in videos of socially interacting flies, was utilized as the time criteria. Here, the time criteria were acquired by increasing the number of bins on the normalized time frequency distribution until the mean of the enclosed bins decreases.

For each species, this algorithm was performed 500 times where 15 videos were randomly sampled with replacement in order to generate a 95% confidence interval and median estimate of each species’ interaction criteria. Both the pilot data and experimental data were analyzed with this algorithm. For a few species, the estimated interaction criteria produced questionable results that contradicted our own observations. The following changes were made to amend the computerized estimates:

1. *D. mauritiana* males: the angle criterion was relaxed to 121.5 degrees since that angle encompasses more hotspots in their social space heat map (Figure S9).
2. *D. santomea* females: the distance was restricted to the lower limit of the 95% confidence interval, the angle was relaxed to the upper limit of the 95% confidence interval, and time was restricted to the lower limit of the 95% confidence interval (Figure S10).
3. *D. immigrans* males: The automated time criterion was estimated to be 0 s, clearly a failure of the algorithm to provide a meaningful criterion. As a result, 103 social interactions of male *D. immigrans* were randomly flagged throughout the videos and the duration of these interactions (in units of seconds) was recorded in a histogram. The minimum value of 0.53 s was used to generate SINs. This value falls in a bin that contains ~ 10% of the interactions observed, making it a reliable approximation of a minimum social interaction time criterion (Figure S11).
4. *D. mojavensis* males and females: the angle criterion computed in the experimental data was narrower than the angle criterion estimated from the pilot data. Because the pilot data angle criterion encompasses a larger distribution of hot spots that are consistent with the experimental data heat maps, the pilot data angle criterion was used for SIN analysis in the experimental data (Figure S12).

Geographic coordinate acquisition for the 20 species

To test the influence of climate on species variation in SINs, climate data was acquired from each species geographic distribution. Many of the species purchased from the *Drosophila* Species Stock Centre had records of the geographic coordinates, and/or the city and country the stock was collected from. For the stocks that did not have precise coordinates listed on the stock center, the rough latitude and longitude of the recorded locations were noted from a search on Google Maps. The coordinates estimating each species’ stock geographic origin are listed in Table S1. However, we were unsure of the precise geographic origin for the following stocks: *D. sechellia*, *D. yakuba*, *D. erecta*, and *D. mojavensis*. As a result, we utilized Taxodros, a database that contains records of coordinates pin-pointing thousands of collection sites for drosophilid species (http://www.taxodros.uzh.ch/). We acquired all the Taxodros coordinates listed for each of our 20 species through custom MATLAB scripts. For the species stocks that had known coordinates from the stock centre or Google Maps, we filtered out all Taxodros coordinates

beyond +/- 2 latitudinal and longitudinal units from the known coordinates (Table S1). Fortunately, the species without known coordinates (*D. mojavensis*, *D. yakuba*, *D. erecta*, *D. sechellia*), have narrow geographic distributions confined to one continent. Therefore, the mean latitude and longitude were calculated from all Taxodros coordinates to approximate the center of their geographic distribution, and these coordinates are listed in Table S1.

The Taxodros database also aided in acquiring more accurate estimates of the climate variables used in the environmental models (Figure 5). Rather than estimating a single measure for all 19 climate variables using the coordinates in Table S1, we acquired distributions for the 19 climate variables from the filtered Taxodros coordinates. For all species, the mean value was calculated from the distribution of each climate variable, and these mean measures were incorporated into the principal component analysis. The first 5 principal components accounted for 92% of the variance across all the climate data (Figure S7). Temperature annual range (BIO7), mean temperature of coldest quarter (BIO11), and precipitation of wettest quarter (BIO16) contribute the most to the variance of these principal components (Figure S8).

Supplementary environmental, behavioural and combined regression models

We characterized the behavioural elements in a hierarchy, where movement is level one, social spacing is level two, pairwise interactions are level three, and SINs are level four (Figure 1). Separate environmental models, behavioural models and combined models were also generated using level two and three variables as response variables. To generate level three behavioural models, level one and two variables (movement, social spacing) served as predictors (Figure S3). To generate level two behavioural models, the level one variable (movement) served as a predictor (Figure S4).

Supplementary Tables and Figures

Table S1-Related to Method Details: Summary of species stocks used for all experiments. All stocks purchased from the Drosophila Species Stock Centre are indicated. Where applicable, the geographic origin of the capture site of each stock is indicated. The approximate geographic coordinates of each stock are also indicated, and these were utilized for acquisition of climate data (see supplementary methods).

Species	**Geographic origin (year captured)**	**GIS coordinates [latitude, longitude]**
D. melanogaster (Canton-S)	Canton, Ohio	[40.46, -81.19]
D. sechellia	Seychelles islands	[-4.15, 55.45]*
D. mauritiana	Mauritius (1987)	[-20.3484, 57.55215]
D. santomea	San Tome and Principe Island (1998)	[0.23, 6.6]
D. yakuba	Central-West Africa	[-1.74361, 18.07164]*
D. erecta	Central-West Africa	[1.863, 3.302]*
D. bipectinata	Cambodia (2011)	[12.56, 104.99]
D. ananassae	Guam (2012)	[13.4443, 144.79373]
D. persimilis	Santa Cruz Island, California (2004)	[37.73, -122.43]
D. willistoni	Jalisco, Mexico (2004)	[20.32,-105.31]
D. paramelanica	Reedsburg, Wisconsin (2009)	[43.5625, -89.8255]
D. melanica	Austin, Texas	[30.27, -97.74]
D. mojavensis	Sonora, Mexico	[29.167, -111.947]*
D. buzzatii	Cochabamba, Bolivia (1958)	[-17.41, -66.16]
D. hydei	Victoria, Australia (2007)	[-37.79, 145.43]
D. novamexicana	Moab, Utah (1949)	[38.57, -109.54]
D. americana	Iowa River, Iowa (2004)	[41.96, -93.89]
D. virilis	Puebla, Mexico (1947)	[19.04, -98.20]
D. immigrans	San Diego, California (2009)	[32.84, -117.20]
Ch. procnemis	Fukuoka, Japan (1981)	[33.58, 130.4]

* Indicates the precise geographic origin of the stock was uncertain

Table S2-Related to Method Details: Interaction criteria of all 20 species for both male and female flies. Interaction criteria in bold are values that were used to override the automated estimations (see methods).

	Male			Female		
Species	**Distance**	**Angle**	**Time**	**Distance**	**Angle**	**Time**
D. ananassae	1.75	140	0.95	2.25	100	1.15
D. bipectinata	2	127.5	1.025	1.75	130	0.7
D. melanogaster	1.75	105	0.45	1.75	110	0.55
D. erecta	2.5	75	0.5	2.25	30	0.85
D. mauritiana	1.75	**121.5**	0.25	2.5	95	1.25
D. persimilis	2	60	0.6	2	110	0.85
D. santomea	1.5	20	0.2	**2.75**	**135**	**0.4**
D. sechellia	1.75	70	0.35	2	85	0.6
D. willistoni	2	70	0.6	2.25	120	1.05
D. yakuba	2.25	55	0.5	2.5	120	1.4
D. americana	1.75	130	2.65	1.5	125	1.85
D. buzzatii	1.25	50	0.95	1.25	115	1.3
D. hydei	1.5	125	0.6	1.75	135	1.3
D. immigrans	1.5	125	**0.53**	1.5	120	1.85
D. melanica	1.25	135	0.9	1.25	130	1
D. mojavensis	1.5	**147.5**	1.3	1.5	**140**	0.35
D. novamexicana	1.25	115	1.3	1.25	120	0.6
D. paramelanica	1.5	130	1.75	1.5	135	1.6
Ch. procnemis	3.25	35	1	2.25	60	0.8
D. virilis	1.25	135	1.75	1.25	140	1.8

Table S3-Related to Figure 4: Leg lengths of all 20 species for male flies. Mean front, middle, and rear legs were measured and a total mean leg length was calculated.

Species	Sample Size	Mean Front Leg (µm)	Mean Middle Leg (µm)	Mean Rear Leg (µm)	Mean Total Leg Length (µm)
Ch. procnemis	20	1808.24	2011.56	1808.24	1876.02
D. melanogaster	10	1743.64	2184.59	2225.42	2051.21
D. sechellia	10	1509.37	1929.62	1978.64	1805.88
D. mauritiana	10	1379.61	1798.25	1911.56	1696.47
D. santomea	10	1358.88	1693.15	1785.28	1612.44
D. yakuba	10	1368.38	1739.29	1805.78	1637.82
D. erecta	10	1352.29	1680.74	1820.83	1617.95
D. bipectinata	10	1326.81	1677.58	1712.40	1572.27
D. ananassae	10	1483.34	1865.13	1892.38	1746.95
D. persimilis	10	1601.28	1893.45	2083.31	1859.35
D. willistoni	10	1403.30	1737.83	1771.76	1637.63
D. paramelanica	10	1852.78	2267.74	2379.89	2166.80
D. melanica	10	1809.51	2266.98	2386.99	2154.49
D. mojavensis	10	1444.44	1678.11	1800.84	1641.13
D. buzzatii	10	1623.25	1952.49	2015.45	1863.73
D. hydei	10	2177.33	2671.00	2772.14	2540.16
D. novamexicana	10	1930.75	2380.17	2529.14	2280.02
D. americana	10	2042.64	2501.99	2657.44	2400.69
D. virilis	10	2172.87	2637.23	2910.11	2573.40
D. immigrans	10	2357.48	2790.73	2893.47	2680.56

Table S4-Related to Figure 4: Mean body size of all 20 species for male flies. Body sizes were calculated from tracked videos and averaged for each species.

Species	Sample Size	Mean Body Size (µm)
Ch. procnemis	20	758.85
D. melanogaster	46	625.95
D. sechellia	22	557.24
D. mauritiana	26	540.98
D. santomea	23	491.66
D. yakuba	24	502.41
D. erecta	25	489.90
D. bipectinata	22	465.98
D. ananassae	23	514.88
D. persimilis	25	634.13
D. willistoni	23	522.65
D. paramelanica	23	734.58
D. melanica	22	749.40
D. mojavensis	25	625.39
D. buzzatii	27	767.20
D. hydei	25	856.84
D. novamexicana	24	860.91
D. americana	27	889.59
D. virilis	25	866.36
D. immigrans	23	791.03

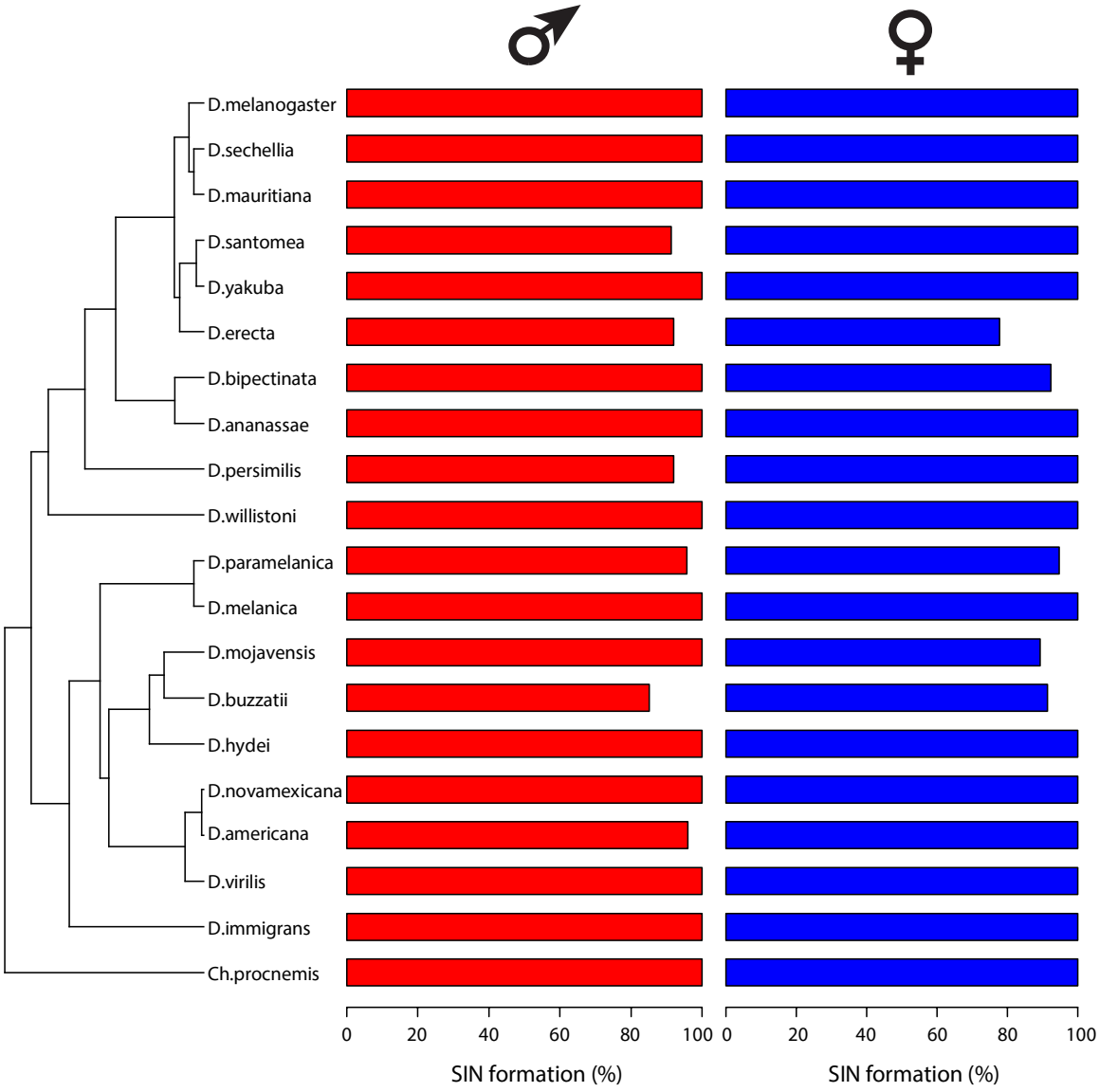
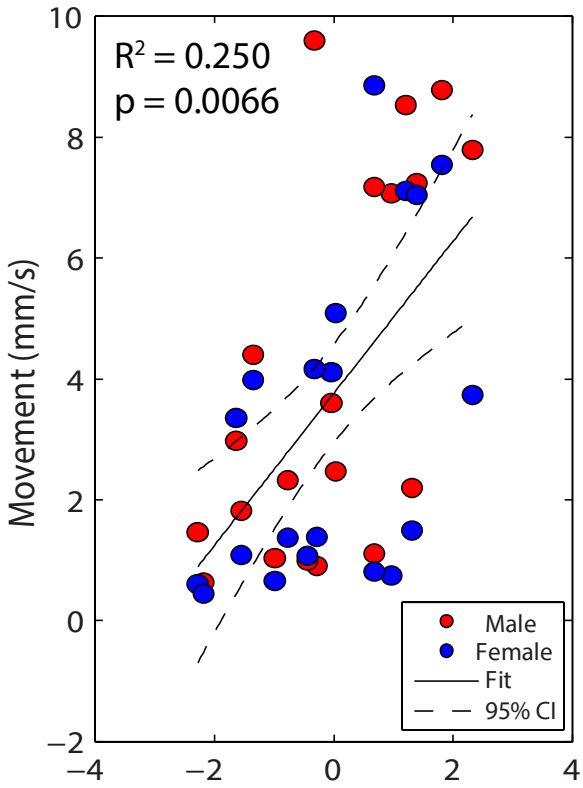


Figure S1-Related to Method Details and Figure 3: Ability to form SINs for the male dataset (red) and female dataset (blue). The x-axis is expressed as a percentage of the number of videos with at least one SIN iteration divided by the total number of videos acquired for each species. There are no differences across species in their ability to form networks ($p = 1$; $\chi 2$ goodness of fit test). The total number of videos that formed SINs are as follows: *D. melanogaster*: n= 46 (male), n=48 (female); *D.sechellia*: n=22 (male), n=22 (female); *D. mauritiana*: n=26 (male), n=27 (female); *D. santomea*: n=21 (male), n=23 (female); *D. yakuba*: n=24 (male), n=26 (female); *D.erecta*: n=23 (male), n=21 (female); *D. bipectinata*: n=22 (male), n=24 (female); *D. ananassae*: n=23 (male), n=25 (female); *D. persimilis*: n=23 (male), n=25 (female); *D. willistoni*: n=23 (male), n=24 (female); *D.paramelanica*: n=22 (male), n=18 (female); *D. melanica*: n=22 (male), n=21 (female); *D. mojavensis*: n=25 (male), n=25 (female); *D. buzzatii*: n=23 (male), n=21 (female); *D. hydei*: n=25 (male), n=24 (female); *D. novamexicana*: n=24 (male), n=22 (female); *D. virilis*: n=25 (male), n=27 (female); *D. immigrans*: n=23 (male), n=22 (female); *Ch. procnemis*: n=20 (male), n=21 (female).



y = 3.76 + 0.43*[PC2] - 0.45*[PC3] + 0.62*[PC4] - 0.9*[PC5]

Figure S2-Related to Figure 5: Environmental model for the level 1 variable Movement. Each data point represents the mean SIN measure for a single species. The mean SIN measure for groups of male flies and female flies were pooled into each regression and are labeled with red and blue points, respectively. The solid trend line indicates the line of best fit and dashed lines indicate 95% confidence interval of the model. The equation for this model is listed below the x-axis.

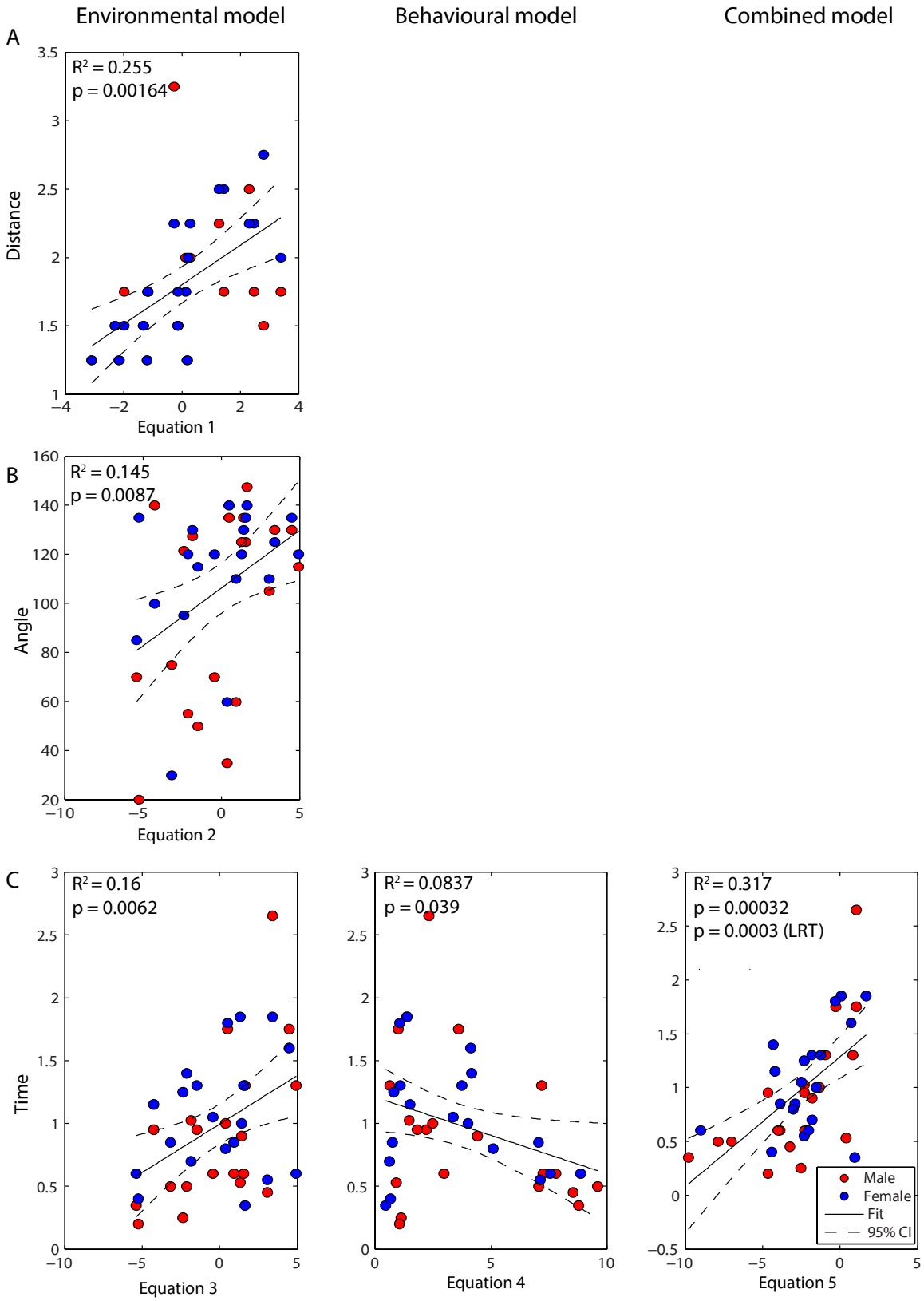

Environmental model
Behavioural model
Combined model
A
B
C
R² = 0.255
p = 0.00164
Distance
Equation 1
R² = 0.145
p = 0.0087
Angle
Equation 2
R² = 0.16
p = 0.0062
Time
Equation 3
R² = 0.0837
p = 0.039
Equation 4
R² = 0.317
p = 0.00032
p = 0.0003 (LRT)
Male
Female
Fit
95% CI
Equation 5

Figure S3-Related to Figure 5: Environmental, behavioural and combined models for the level 2 social spacing variables: Distance, Angle and Time. For all regressions, each data point represents the mean SIN measure for a single species. The mean SIN measure for groups of male flies and female flies were pooled into each regression and are labeled with red and blue points, respectively. Each solid trend line indicates line of best fit and dashed lines indicate 95% confidence interval of the model. **A)** No significant behavioural model formed for distance, therefore no combined model to report. The environmental model (Equation 1 predicts social distance (R^2 = 0.255, p = 0.00164). **B)** No significant behavioural model formed for angle, therefore no combined model to report. The environmental model (Equation 2) predicts angle (R^2 = 0.145, p = 0.0087). **C)** The environmental model (Equation 3) predicts time (R^2 = 0.16, p = 0.0062). The behavioural model (Equation 4) predicts time (R^2 = 0.0837, p = 0.039). The combined model significantly improves prediction of time compared to the behavioural model alone (p = 0.0003, likelihood ratio test). **Equation 1:** y = 1.80 – 0.071*[PC1] – 0.13*[PC5]. **Equation 2:** y = 106.3 + 4.71*[PC1]. **Equation 3:** y = 0.99 + 0.077*[PC1]. **Equation 4:** y = 1.21 – 0.061*[movement]. **Equation 5:** y = 1.28 + 0.091*[PC1] – 0.078*[movement].

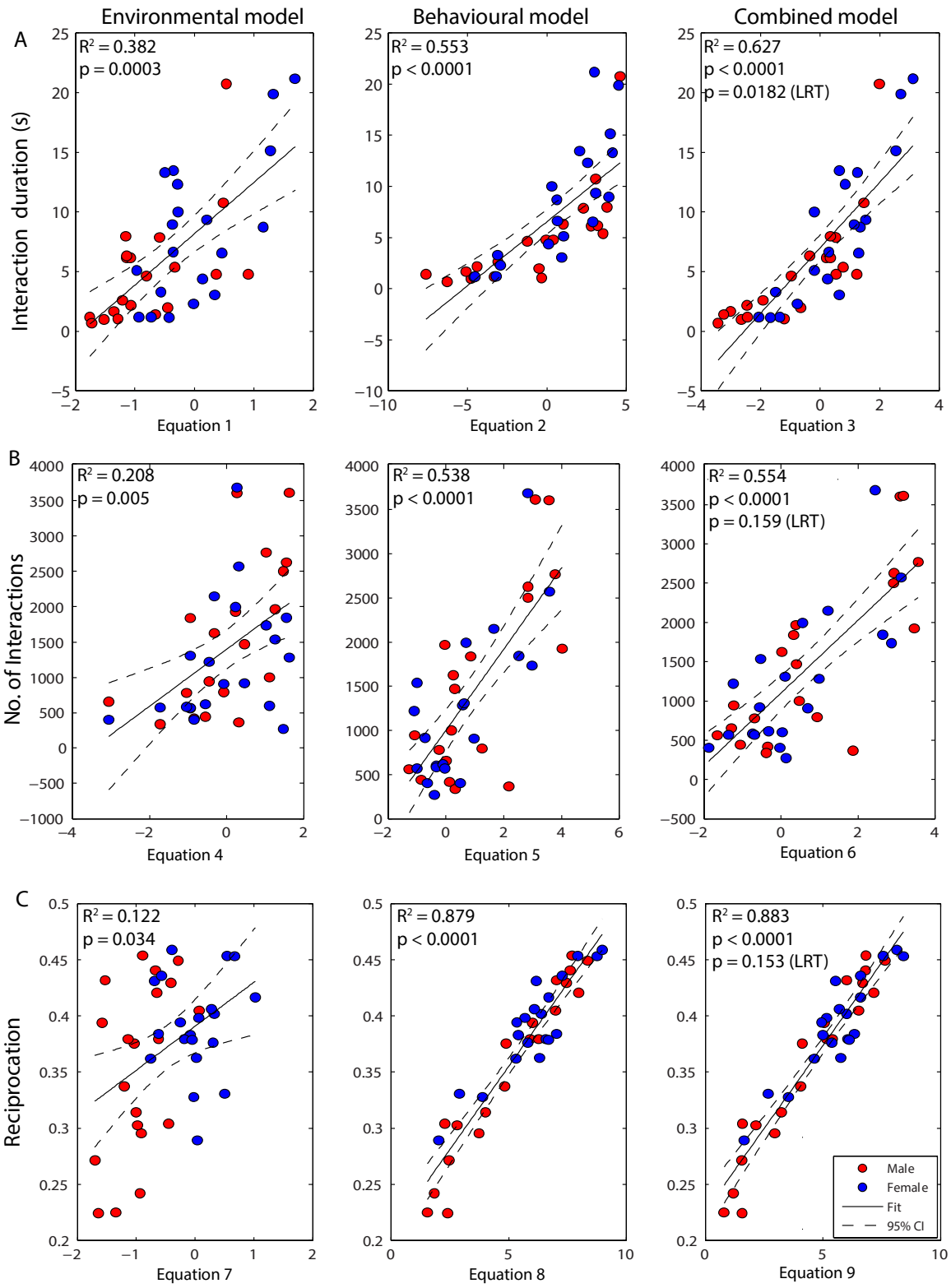

Environmental model
Behavioural model
Combined model
A
B
C
Interaction duration (s)
No. of Interactions
Reciprocation
R² = 0.382
p = 0.0003
R² = 0.553
p < 0.0001
R² = 0.627
p < 0.0001
p = 0.0182 (LRT)
R² = 0.208
p = 0.005
R² = 0.538
p < 0.0001
R² = 0.554
p < 0.0001
p = 0.159 (LRT)
R² = 0.122
p = 0.034
R² = 0.879
p < 0.0001
R² = 0.883
p < 0.0001
p = 0.153 (LRT)
Equation 1
Equation 2
Equation 3
Equation 4
Equation 5
Equation 6
Equation 7
Equation 8
Equation 9
Male
Female
Fit
95% CI

Figure S4-Related to Figure 5: Environmental, behavioural and combined models for the level 3 variables interaction duration, number of interactions and reciprocation. For all regressions, each data point represents the mean SIN measure for a single species. The mean SIN measure for groups of male flies and female flies were pooled into each regression and are labeled with red and blue points, respectively. Each solid trend line indicates line of best fit and dashed lines indicate 95% confidence interval of the model. **A)** The environmental model (Equation 1) predicts interaction duration (R^2 = 0.382, p = 0.0003). The behavioural model (Equation 2) predicts interaction duration (R^2 = 0.553, p < 0.0001). The combined model (Equation 3) does not significantly improve the prediction of interaction duration compared to the behavioural model alone (p = 0.0182, likelihood ratio test). **B)** The environmental model (Equation 4) predicts number of interactions (R^2 = 0.20772, p = 0.005). The behavioural model (Equation 5) predicts number of interactions (R^2 = 0.538, p < 0.0001). The combined model does not significantly improve the prediction of number of interactions compared to the behavioural model alone (p = 0.1594, likelihood ratio test). **C)** The environmental model (Equation 7) predicts reciprocation (R^2 = 0.122, p = 0.034). The behavioural model (Equation 8) predicts reciprocation (R^2 = 0.879, p < 0.0001). The combined model (Equation 9) does not significantly improve the prediction of reciprocation compared to the behavioural model alone (p = 0.153, likelihood ratio test). **Equation 1:** y = 8.17 – 0.76*[PC2] + 1.02*[PC3] + 2.31*[PC5]. **Equation 2:** y = 6.54 – 1.25*[movement] + 0.045*[angle]. **Equation 3:** y = 6.99 – 0.23*[PC2] + 0.47*[PC3] + 1.47*[PC5] – 0.93*[movement] + 0.038*[angle]. **Equation 4:** y = 1398 + 220*[PC4] – 336*[PC5]. **Equation 5:** y = 1001 + 213*[movement] – 406*[time]. **Equation 6:** y = 1100 + 91.5*[PC4] – 151*[PC5] + 183*[movement] – 387*[time]. **Equation 7:** y = 0.39 + 0.011*[PC3]. **Equation 8:** y = 0.21 – 0.0065*[movement] + 0.014*[social distance] + 0.0017*[angle] – 0.025*[time]. **Equation 9:** y = 0.22 + 0.0028*[PC3] – 0.0061*[movement] + 0.01*[social distance] + 0.0017*[angle] – 0.025*[time].

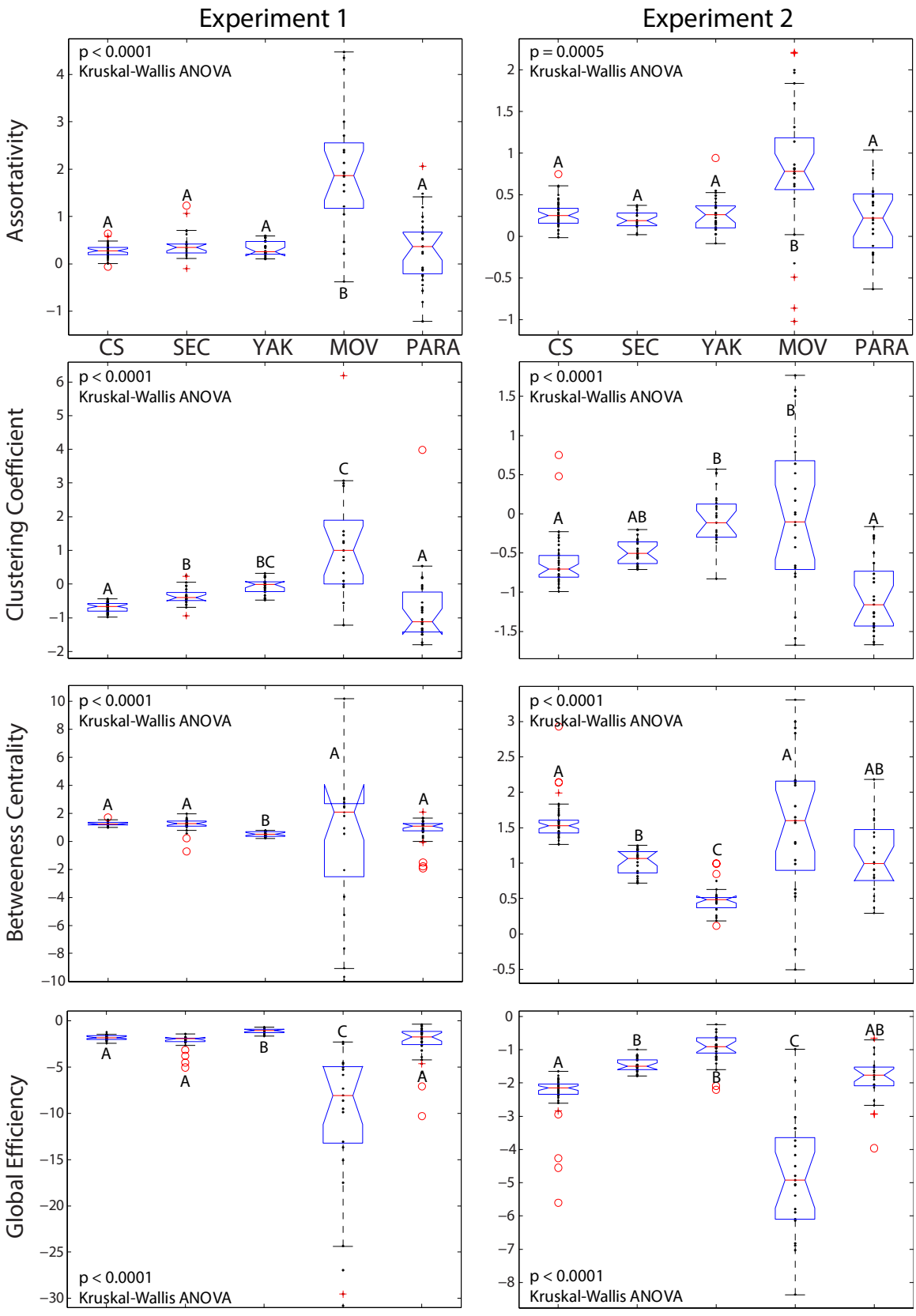

Experiment 1
Experiment 2
Assortativity
Clustering Coefficient
Betweeness Centrality
Global Efficiency
p < 0.0001
Kruskal-Wallis ANOVA
p = 0.0005
Kruskal-Wallis ANOVA
CS
SEC
YAK
MOV
PARA

Figure S5-Related to Figure 3: Relative species differences between *D. melanogaster* (CS), *D. sechellia* (SEC), *D. yakuba* (YAK), *D. mojavensis* (MOV), *D. paramelanica* (PARA) are replicated in male flies between pilot data (Experiment 1; left panels) and experimental data (Experiment 2; right panels). All figures are boxplots which outline the distribution of the z-scores for all four SIN measures. Red circles indicate outliers that were removed prior to statistical analysis. Letters above each box indicate statistically distinct groups from a Kruskal-Wallis one-way ANOVA followed Tukey-Kramer post-hoc tests.

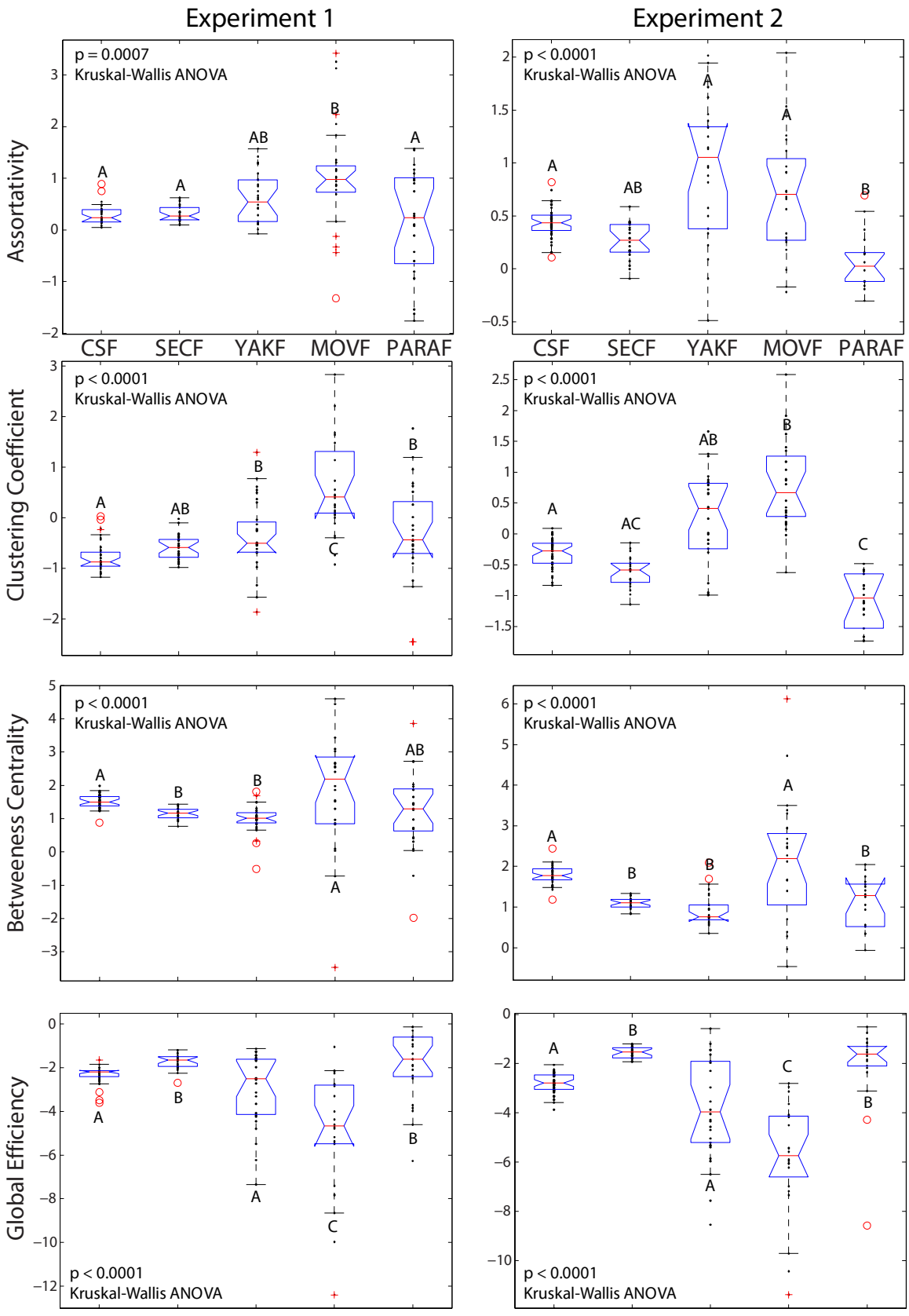

Experiment 1
Experiment 2
Assortativity
Clustering Coefficient
Betweeness Centrality
Global Efficiency
p = 0.0007
Kruskal-Wallis ANOVA
p < 0.0001
Kruskal-Wallis ANOVA
CSF
SECF
YAKF
MOVF
PARAF

Figure S6-Related to Figure 3: Relative species differences between *D. melanogaster* (CS), *D. sechellia* (SEC), *D. yakuba* (YAK), *D. mojavensis* (MOV), *D. paramelanica* (PARA) are replicated in female flies between pilot data (Experiment 1; left panels) and experimental data (Experiment 2; right panels). All figures are boxplots which outline the distribution of the z-scores for all four SIN measures. Red circles indicate outliers that were removed prior to statistical analysis. Letters above each box indicate statistically distinct groups from a Kruskal-Wallis one-way ANOVA followed Tukey-Kramer post-hoc tests.

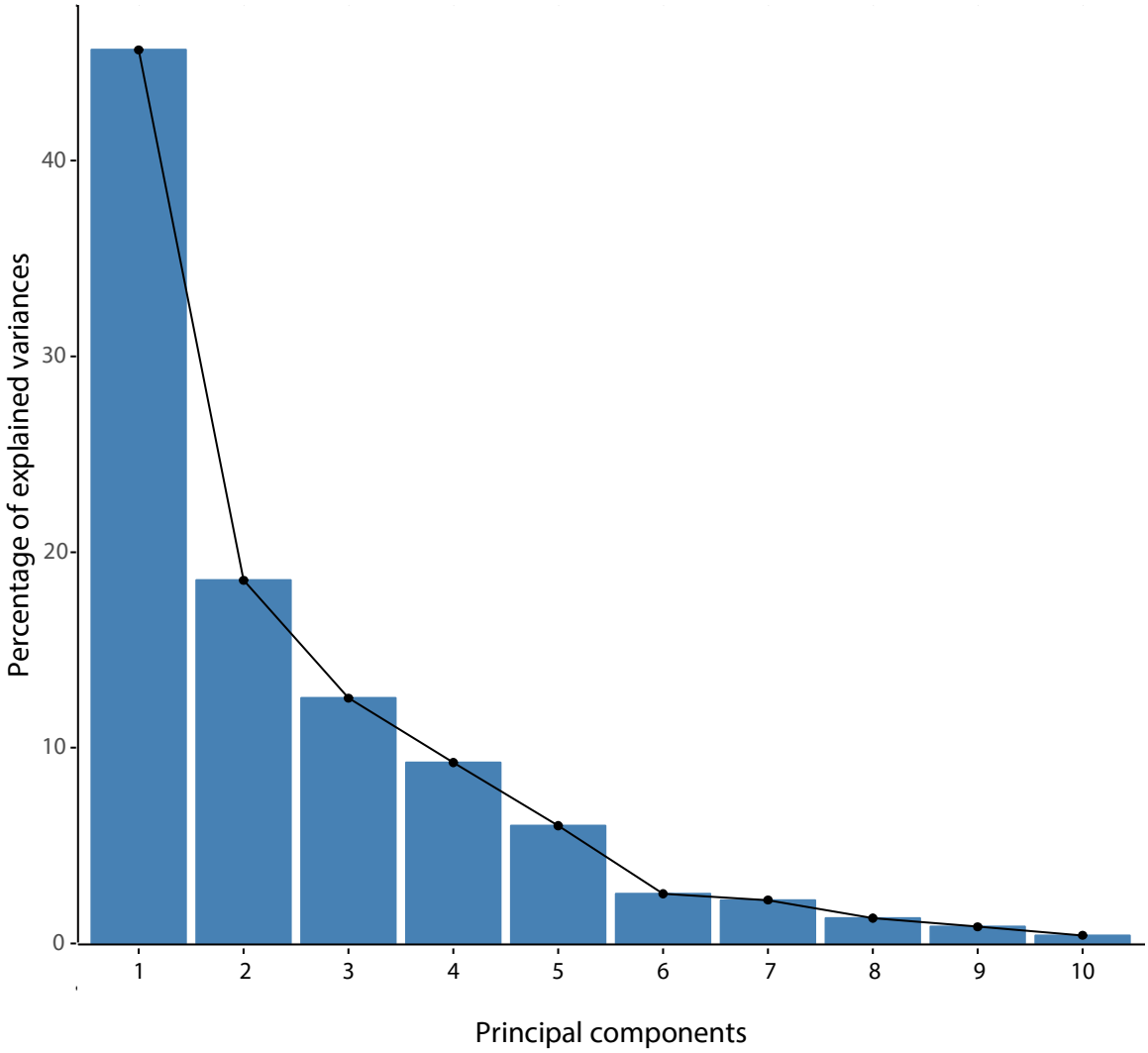
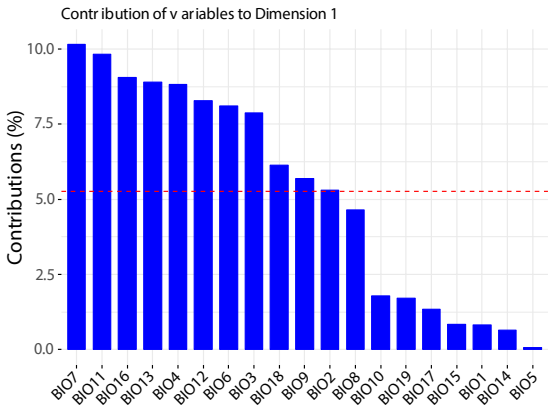
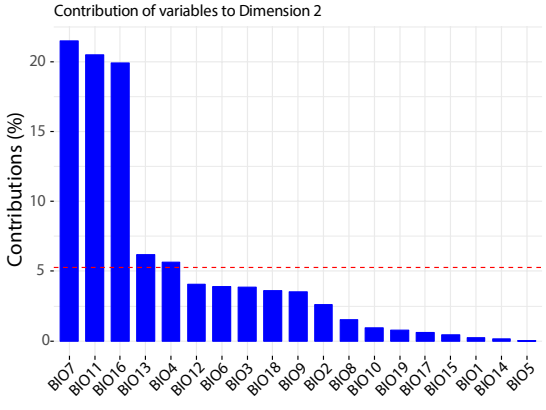
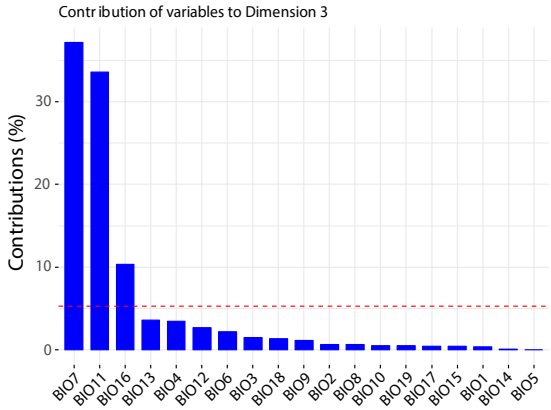
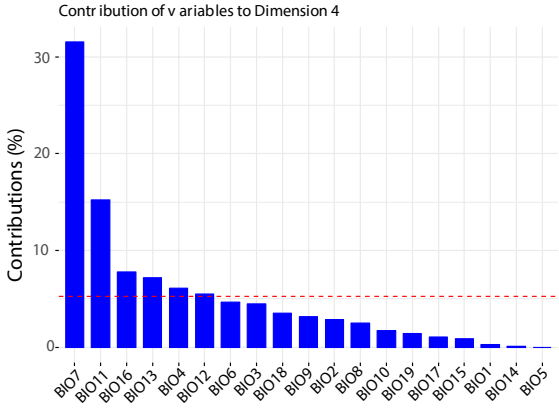


Figure S7-Related to Figure 5: The first 5 principal components of the climatic measures, extracted from each species geographic origin, account for 92% of the variance across the 19 variables. Each bar represents the percentage of variance explained by the first 10 dimensions. Because the rate of decrease reduces after dimension 5, the first 5 dimensions were used to represent the climate variables for all regression analyses.


Contr ibution of v ariables to Dimension 1
Contributions (%)
10.0
7.5
5.0
2.5
0.0
BIO7
BIO11
BIO16
BIO13
BIO4
BIO12
BIO6
BIO3
BIO18
BIO9
BIO2
BIO8
BIO10
BIO19
BIO17
BIO15
BIO1
BIO14
BIO5


Contribution of variables to Dimension 2
Contributions (%)
20
15
10
5
0
BIO7
BIO11
BIO16
BIO13
BIO4
BIO12
BIO6
BIO3
BIO18
BIO9
BIO2
BIO8
BIO10
BIO19
BIO17
BIO15
BIO1
BIO14
BIO5


Contr ibution of variables to Dimension 3
Contributions (%)
30
20
10
0
BIO7
BIO11
BIO16
BIO13
BIO4
BIO12
BIO6
BIO3
BIO18
BIO9
BIO2
BIO8
BIO10
BIO19
BIO17
BIO15
BIO1
BIO14
BIO5


Contr ibution of v ariables to Dimension 4
Contributions (%)
30
20
10
0
BIO7
BIO11
BIO16
BIO13
BIO4
BIO12
BIO6
BIO3
BIO18
BIO9
BIO2
BIO8
BIO10
BIO19
BIO17
BIO15
BIO1
BIO14
BIO5

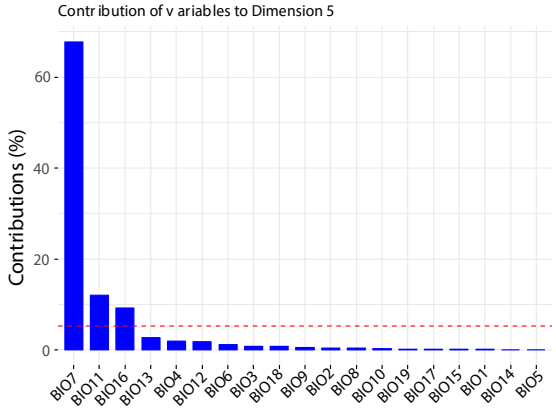

Contr ibution of v ariables to Dimension 5
Contributions (%)
60
40
20
0
BIO7
BIO11
BIO16
BIO13
BIO4
BIO12
BIO6
BIO3
BIO18
BIO9
BIO2
BIO8
BIO10
BIO19
BIO17
BIO15
BIO1
BIO14
BIO5

Figure S8 – Related to Figure 5: The relative contribution of all WorldClim variables towards the variance of each of the 5 principal components. Temperature annual range (BIO7), mean temperature of coldest quarter (BIO11), and precipitation of wettest quarter (BIO16) contribute the most to the variance of each principal component. **BIO1:** annual mean temperature; **BIO2:** mean diurnal range (calculated as mean of monthly(max temp – min temp); **BIO3:** isothermability (calculated as BIO2/BIO7 * 100); **BIO4:** temperature seasonality; **BIO5:** maximum temperature of warmest month; **BIO6:** minimum temperature of coldest month; **BIO7:** temperature annual range (calculated as BIO5-BIO6); **BIO8:** mean temperature of wettest quarter; **BIO9:** mean temperature of driest quarter; **BIO10:** mean temperature of warmest quarter; **BIO11:** mean temperature of coldest quarter; **BIO 12:** annual precipitation; **BIO13:** precipitation of wettest month; **BIO14:** precipitation of driest month; **BIO15:** precipitation seasonality; **BIO16:** precipitation of wettest quarter; **BIO17:** precipitation of driest quarter; **BIO18:** precipitation of warmest quarter; **BIO19:** precipitation of coldest quarter.

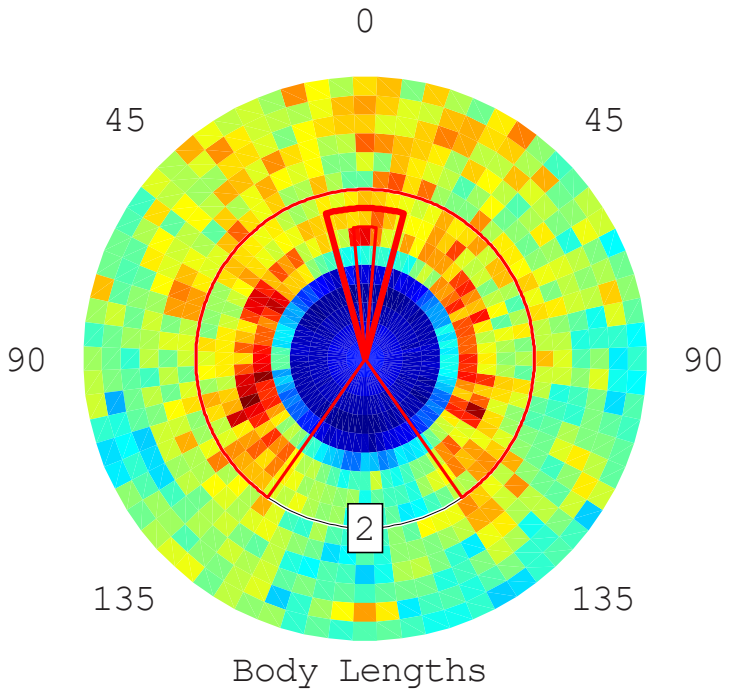


Figure S9: Heat map associated with the social spacing criteria calculated for *D. mauritiana* male flies. This heat map represents the distribution of spatial positions mapped between all interacting flies in all video trials analyzed. Red regions represent interaction hot spots and blue regions represent interaction cold spots. The bolded red outline represents the median distance and angle criteria (distance = 1.75 body lengths; angle = 15 degrees) while the unbolded outlines represent the upper and lower limits of the 95% confidence interval obtained from the bootstrapping analysis (distance: 1.5-2 body lengths; angle; 5-145 degrees). Since many interaction hotspots are evident at larger angles, the angle criterion applied to SIN analysis was relaxed to 121.5 degrees.

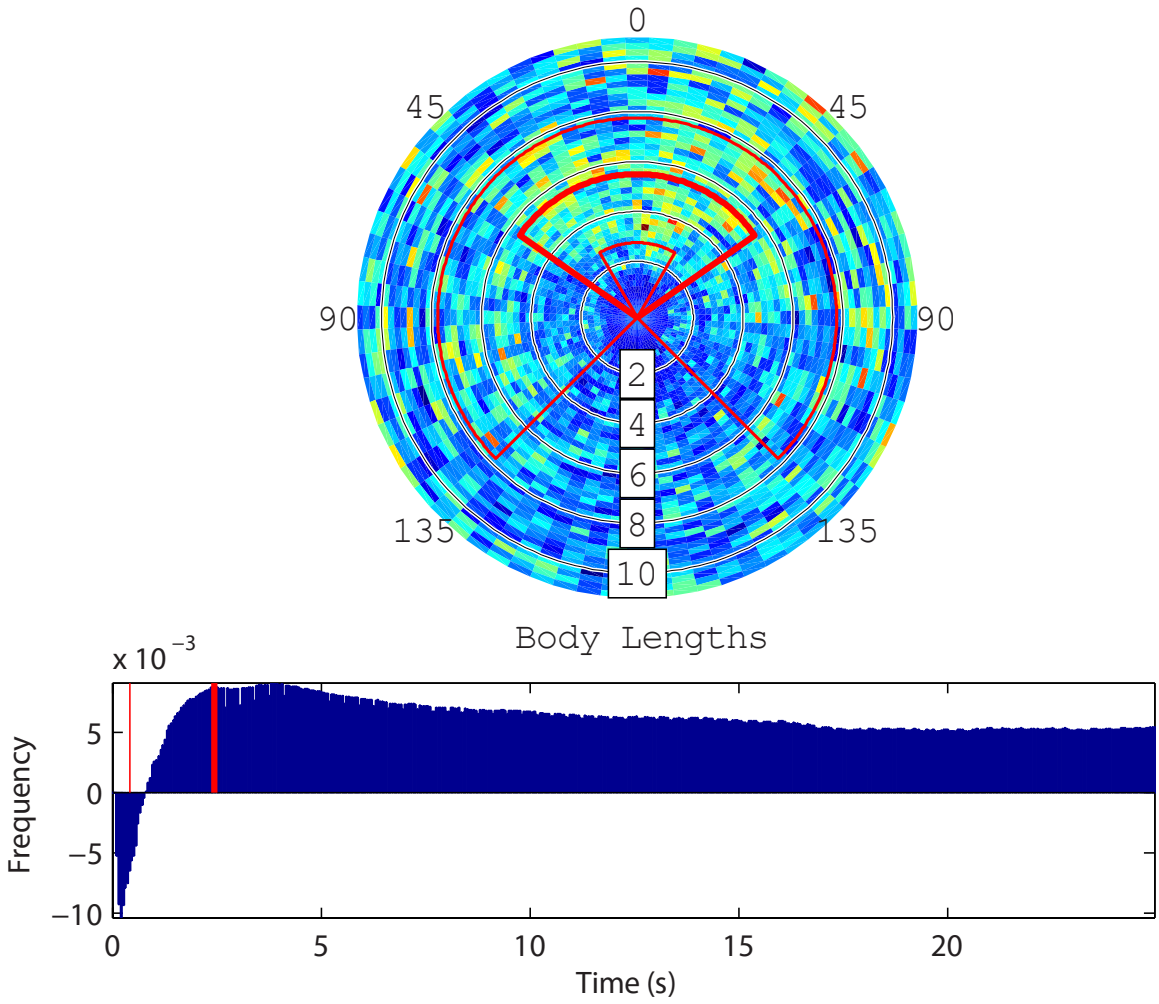


Figure S10: Heat map and histogram associated with the social spacing criteria calculated for *D. santomea* female flies. This heat map represents the distribution of spatial positions mapped between all interacting flies in all video trials analyzed. Red regions represent interaction hot spots and blue regions represent interaction cold spots. The bolded red outline represents the median distance and angle criteria (distance = 5.5 body lengths; angle = 55 degrees) while the unbolded outlines represent the upper and lower limits of the 95% confidence interval obtained from the bootstrapping analysis (distance: 2.75-7.75 body lengths; angle; 30-135 degrees). The histogram below the heat map represents the distribution of the time flies spend interacting using the distance and angle criteria extrapolated from the heatmap. Positive bins represent time criteria over-represented in the video trials whereas negative bins represent time criteria over-represented in the virtual trials. The bolded red line represents the median time criteria (2.425 seconds) while the unbolded red lines indicate the 95% confidence interval (0.4-63.1 seconds) estimated from bootstrapping. The sparse interaction hotspots in the heatmap likely resulted from the sedentary nature of this species, producing questionable social spacing criteria.

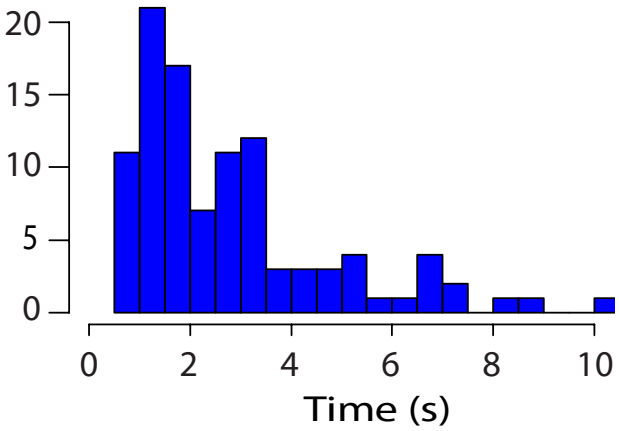


Figure S11: Histogram representing the distribution of the interaction duration of 103 randomly sampled sequences of social interactions in *D. immigrans* male flies across 23 video trials. Since the automated time criteria for *D. immigrans* was estimated to be 0 seconds, we relied on an alternative estimate by observing randomly sampled video sequences. The minimum value in this distribution (0.53 seconds) falls in a bin that contains ~ 10% of the interactions observed, making it a reliable approximation of a minimum social interaction time criterion.

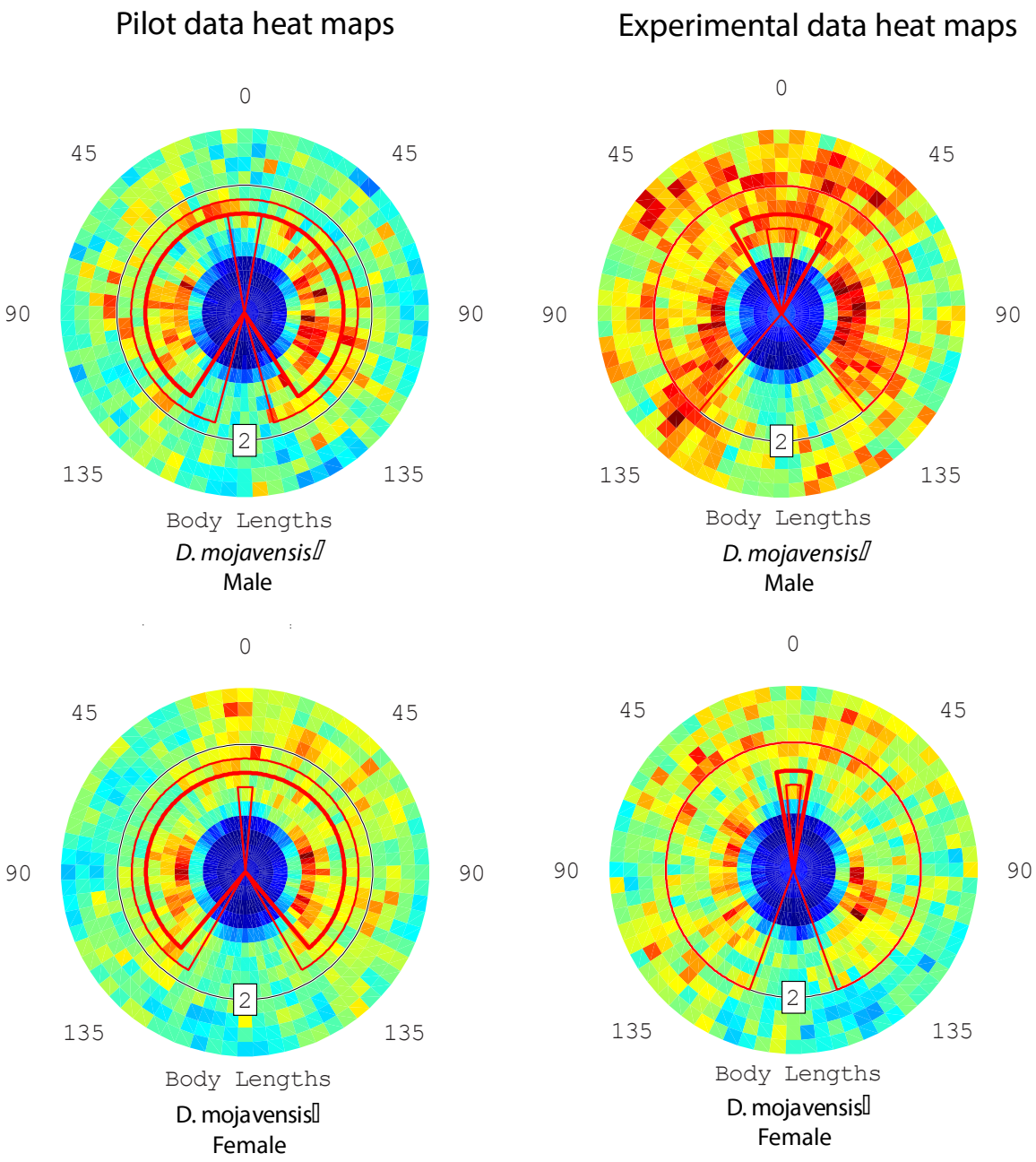


Figure S12: Heat maps associated with the social spacing criteria for *D. mojavensis* males and females in the pilot data (left) and experimental data (right). Heat maps are consistent between the pilot and experimental data. However the estimated angle criterion in the experimental data (male: median = 30 degrees, 95% CI = 10-140; female: median = 10 degrees, 95% CI = 5-160 degrees) does not reflect the wide range of interaction hot spots seen in the heatmap. As a result, the angle criterion estimated from the pilot data (male: 147.5 degrees; female: 140 degrees) was used for SIN analysis in the experimental data.

SI References

1. T. A. Markow, P. M. O'Grady, *Drosophila: A guide to species identification and use* (Elsevier, Inc, San Diego, California, USA, 2006), pp. 254.
2. J. Schneider, M. H. Dickinson, J. D. Levine, Social structures depend on innate determinants and chemosensory processing in Drosophila. *Proc Natl Acad Sci U S A* **109 Suppl 2**, 17174-17179 (2012).
3. S. P. Blomberg, T. Garland, A. R. Ives, Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717-745 (2003).
4. A. R. Ives, P. E. Midford, T. Garland Jr, Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology* **56**, 252-270 (2007).
5. J. Schneider, J. D. Levine, Automated identification of social interaction criteria in Drosophila melanogaster. *Biol Lett* **10**, 20140749 (2014).
6. M. Newman, *Networks: an introduction* (Oxford University Press, 2010).