1 **Quantifying between-individual variation using high-throughput phenotyping of**

2 **behavioural traits in the fruit fly (*Drosophila melanogaster*)**

3 Erin L. Macartney[1], Patrice Pottier[1], Samantha Burke[1], Shinichi Nakagawa*[1], Szymon M.

4 Drobniak*[1,2]

5 * - shared senior authorship

6

7 [1] Evolution and Ecology Research Centre, School of Biological Earth and Environmental

8 Science, University of New South Wales, Sydney, Australia

9 [2] Institute of Environmental Sciences, Jagiellonian University, Krakow, Poland

10

11 Corresponding authors: s.drobniak@unsw.edu.au & erin.macartney@unsw.edu.au

12 Erin L. Macartney (ORCiD: 0000-0003-3866-143X, Twitter: @Erin_Macartney_)

13 Patrice Pottier (ORCiD: 0000-0003-2106-6597, Twitter: @Patrice_Science)

14 Samantha Burke (ORCiD: 0000-0001-6902-974X)

15 Shinichi Nakagawa (ORCiD: 0000-0002-7765-5182, Twitter: @itchyshin)

16 Szymon M. Drobniak (ORCiD: 0000-0001-8101-6247, Twitter: @szymekdr)

17

18

19 **Abstract**

20 Behavioural phenotyping is often time and labour-intensive, which can come at a cost to
21 sample size and statistical precision. This is particularly a concern given that behaviours are
22 often highly variable within and between individuals, so naturally requires a larger sample
23 size. *Drosophila melanogaster* is a common model system in many research fields, and
24 behavioural observations are frequently required. While *D. melanogaster* has a rapid
25 lifecycle that enables large numbers of flies to be reared for experiments, they are still
26 subject to methodological bottlenecks for behavioural observations. Additionally, their small
27 and delicate bodies make it difficult to observe certain behaviours in real-time, for example,
28 in movement tracking or when performing repeated assays on the same individuals. Here,
29 we present a method, pilot data, custom data processing and analysis scripts for high-
30 throughput behavioural phenotyping in *D. melanogaster*, as well as general remarks for
31 future studies. We used automatic tracking units to measure three behaviours in the same
32 individuals: locomotor activity, exploratory behaviour in a Y-maze, and habituation to a
33 startle response stimulus. We then examined between-individual variation and trait
34 correlations using our pilot data. Through this, we show that these behaviours are amenable
35 to high-throughput automated tracking, with locomotor activity generating the most
36 straightforward and high-quality data. These methods can be used to free up time and
37 labour to allocate to increasing sample sizes and can be used to address a range of biological
38 questions in ecology, evolution, and beyond.

39

40 **Key words:** locomotion, startle, habituation, variation, power, precision, automated
41 phenotyping, repeatability

42

43 **Introduction**

44 Behaviour is closely tied to how an individual receives and processes information, and thus,
45 how it responds to its environment [1]. It is also intricately linked to fitness and can represent
46 individual condition and age [2]. Measuring behavioural traits among individuals provides key
47 information regarding population averages and variation while measuring behaviour within
48 individuals allows researchers to quantify behavioural repeatability [3], often referred to as
49 'personality' [4,5]. Furthermore, measuring multiple behaviours within individuals allows the
50 quantification of between-individual behavioural correlations [6], or behavioural syndromes [7].
51 Therefore, accurately measuring behaviour is a key part of individual phenotyping with
52 broad implications across many fields of biology.

53 In contrast to genotyping and common morphological phenotyping, quantifying behaviour
54 can be highly laborious and time-consuming. Measurements of individual behaviour
55 regularly involve directly observing (either in real-time or through video recordings) and
56 quantifying behaviours across time [8–11]. However, this often results in methodological
57 bottlenecks (e.g., an inability to measure enough individuals in the most
58 appropriate/physiologically relevant time-window) and can come at a cost to sample size [12].
59 Moreover, experimental designs aimed at increasing sample size often inadvertently involve
60 the addition of unwanted variation by using small-sample size experimental blocks. While
61 the use of randomized blocks can be a highly effective experimental design that can enable
62 researchers to reach appropriate sample sizes [13], certain conditions (e.g., temperature,
63 lighting, time of day, unintentional use of different genetic backgrounds or clutches) are not
64 always able to be completely homogenous between blocks, even in laboratory
65 environments. Such compromises that occur within behavioural studies can result in
66 decreased precision of estimates, especially when not accounted for correctly during

67 statistical analysis [14]. This is particularly a concern given that behavioural traits are often
68 highly variable within and between individuals and thus, naturally require larger sample
69 sizes [15,16].

70 The fruit fly, *Drosophila melanogaster* is a common model system that is used to address a
71 range of research questions, including those related to behaviour. For example, fruit flies
72 are common in studies of mating and courtship behaviour [8,17,18], aggression [10,19], and
73 learning and memory [20–22] using a variety of behavioural assays and responses. The fast life-
74 history, short generation time, and ease of breeding hundreds of individuals at once are key
75 benefits of using *D. melanogaster* in behavioural studies. However, quantifying behaviour is
76 still subject to procedural bottlenecks due to the time and labour required to observe
77 individual flies directly. An added difficulty lies in the small size and delicate body structure
78 of fruit flies, making repeated handling of the same individuals, as well as large numbers of
79 flies in a short time, challenging. One solution to such bottlenecks is automated
80 phenotyping. While not all behaviours are easily amenable to automation (e.g., mating
81 behaviours where relevant movements are often subtle and subject to nuances that may be
82 hard to detect through automation, but see, Dankert et al, 2009[23]), some behaviours are likely
83 to be easier to quantify through automation, thus easing the bottleneck and freeing up time
84 and labour that can be invested in increasing sample size and statistical power.

85 Here, we present a method (including protocol, pilot test data, and analysis scripts) for
86 quantifying between-individual variation in behaviour in *D. melanogaster* using automated
87 high-throughput phenotyping. We employed commercially available phenotyping units (see
88 methods) that can perform high-throughput automated behavioural phenotyping in a range
89 of model species, including *D. melanogaster*. Accompanying our detailed methods, we

90 provide test data on three different behaviours measured at an individual level: overall
91 locomotor activity, exploratory behaviour in a Y-maze, and habituation to a startle response
92 stimulus. Through automation, we were able to assay all three behaviours in 360 individuals
93 across four days – a sample size that would not have been easily attainable in this period
94 without a large team of full-time researchers. Furthermore, these behaviours occur very
95 rapidly and thus require specialised recording tools to provide sufficient sensitivity and
96 robustness. Therefore, we show that these methods can be used to assay multiple
97 behaviours in the same individuals, enabling the quantification of between-individual
98 variation in multiple behaviours, as well as examining the relationship between these
99 behaviours (e.g., to determine behavioural syndromes).

100 **Methods**

101 *Fly husbandry*

102 Our study used a large, outbred population of Canton-S wild-type fruit flies with overlapping
103 generations maintained at the University of New South Wales, Sydney. The flies were under
104 the 12h:12h photoperiod in a temperature-controlled room maintaining constant thermal
105 conditions (25°C) and humidity of approximately 45-65%. The colony was kept in a 10-litre
106 plastic terrarium with 12 bottles of commercially available food medium (Carolina Biological
107 Supply, Burlington, NC). The three oldest bottles of food were replaced with three bottles of
108 fresh food once a week.

109 To obtain test individuals, we removed the three newest bottles of food after 24 hours from
110 the stock population and kept each bottle individually until the flies eclosed. This was done
111 to approximately standardise larval density and the age of the individuals to be used in the
112 test. The flies were then aged for six days post-eclosion before being used in the assays.

113 Before each batch of assays, a sample of several dozen flies were transferred into an empty,
114 dry culture bottle and briefly (3-5 minutes) submerged in a bucket of ice to induce a chilling
115 coma. Following coma induction, the flies were transferred into their assay arenas (see
116 Behavioural assays) using aspirators. Before each assay, the flies were allowed to return to
117 full activity for approximately 15 minutes by leaving a loaded set of arenas in a lit 25°C area.
118 Behavioural assays were performed in the mornings and afternoons to avoid the mid-day
119 low-activity period.

120 Each batch of tested flies consisted of individuals repeatedly tested in three different
121 behavioural assays (see Behavioural assays; Fig. 1). Transferring the flies between assays
122 using different arena designs was achieved by briefly inducing chilling comas by placing the
123 arenas into a -20°C freezer until anesthetised. They were then aspirated into the new arena
124 (i.e., between the 48 well plates or y-mazes; see Behavioural assays). Following all tests, the
125 flies were euthanised by leaving them overnight in a freezer (-20°C). The following day all
126 individual insects were sexed. In total, we assayed 360 flies arranged into 8 batches (i.e.,
127 four days of morning and afternoon batches) of 45 flies each (90 flies per day) (see General
128 remarks on how sample size can be increased further).

129

130 *Behavioural assays*

131 We employed three types of assays, each performed using a multi-well plexiglass or
132 polypropylene plate able to hold between 15 and 48 flies (depending on the assay type; Fig.
133 1). Both male and female flies were randomly allocated within the plates and identities were
134 tracked across all assays so that we could later link individual behaviours measured in the

135 same individual. Tracking was achieved by maintaining a strict correspondence between
136 arenas in subsequent assays.

137 *Locomotion tracking:* In this assay, individual flies were loaded into 1 cm deep, round,
138 transparent arenas arranged into a 48-well plate. Because one of the assays (the Y-maze)
139 provided space to test 45 flies (3 plates of 15 flies each), the 48-well plates were always
140 filled with 45 flies whilst leaving the last three arenas empty. For the purposes of this study,
141 we have used 48-well cell-culture plates (NEST Biotechnology Co., Ltd.; China); the choice of
142 this brand was dictated by the negligible gap between the well rim and lid, which prevents
143 flies from escaping their allocated well and migrating to adjacent wells. Locomotion of
144 individual flies was followed by recording their position (see the Phenotyping units section
145 for technical details) in short intervals, which was then used to calculate the distance
146 travelled by each fly in a set time interval. In our assay, the flies were tracked for 5
147 consecutive intervals, 10 minutes each (a total of 50-minutes observation).

148 *Light-off startle assay*: In this assay, we have used the established startle response of fruit
149 flies in response to a brief light-off stimulus [24,25]. The response is elicited by a short 'light-off'
150 pulse lasting 15 milliseconds (ms) and can be measured as a sudden and very fast
151 movement (often involving flight) of the flies. The stimulus was delivered by an optogenetic
152 light-conducting plate mounted beneath the 48-well plate (the same plates that were used
153 for the locomotion assays). The device (part of the ZANTIKS MWP phenotyping system, see
154 the 'Phenotyping units' section for more details) consists of a plexiglass plate coupled with a
155 set of several bright LEDs. The diodes emit green light (530 nm), flooding the experimental
156 arenas with over 7000 lux of light. The system is connected to the phenotyping unit and
157 delivers stimuli in the form of brief light-off pulses.

158 The objective of our experiment was to study habituation to the light-off stimulus. Thus, in
159 our assay, we subjected the flies to three consecutive 15-ms light-off pulses, 1 second apart.
160 The startle response (distance covered by each fly) was recorded within a 1-second interval
161 following each light-off pulse.

162 *Y-maze tracking*: The third assay was performed to explore *D. melanogaster* behaviour in a
163 simple 3-arm maze [26–29]. The test was performed in a different type of arena: flies were
164 loaded into small plexiglass blocks with a y-shaped forking channel etched inside of them.
165 The maze was covered by a sliding coverslip that allowed for easy loading and unloading of
166 mazes with flies. Each plate contained 15 y-mazes. Tracking of flies' behaviour in the mazes
167 lasted 30 minutes. Apart from recording the time spent by flies in each of the three maze
168 arms we have also recorded each arm crossing event (i.e., a fly crossing from one maze arm
169 to another, through the central "neutral zone").

170 The protocol we used was based on a comparative study looking at Y-maze exploration
171 behaviour in mice, zebrafish, and fruit flies [29]. Briefly, the assay generates a sequence of
172 zone locations visited by flies. Transition between two zones is classified as either left (L) or
173 right (R) turn. Analysis involves creating a sliding window along the sequence of turns
174 executed by each fly and grouping the turns in triplets ('trigrams'). The RRR and LLL trigrams
175 are then classified as full repetitions (i.e., 3 consecutive turns in one direction), and RLR/LRL
176 trigrams are classified as full alternations (sequences of 3 turns, each in the opposite
177 direction to the previous). Proportions of full alternations and full repetitions are calculated
178 as basic indices of Y-maze exploration by dividing respective counts by the total number of
179 recorded turn triplets. Similar analyses involving automated tracking of fruit flies in a non-

180 baited Y-maze have already been performed (e.g., in the context of neuronal control of
181 handedness[27]), but our method offers considerably higher time efficiency and throughput.

182 *Phenotyping units*

183 All behavioural tests were performed using automated tracking units produced by Zantiks
184 (Cambridge, UK). The units we employed were from the WMP series, suitable for tracking of
185 small-sized animals such as small insects, crustaceans, fish larvae, etc. Each unit consists of a
186 computer that controls its operation, an experimental chamber that can host experimental
187 arenas (and can be connected to a temperature control unit, able to maintain internal
188 chamber temperature in a narrow, set range) and a camera system able to track animal
189 movement in arenas inserted to the chamber.

190 All the arenas used in our assays had the same format (i.e., dimensions of a standard ELISA
191 multi-well plate) and were placed inside each unit on a raised stand (locomotion and Y-maze
192 assays) or the designated optogenetic stand (light-off startle assay). To be able to track the
193 animals, the unit requires a correct definition of experimental arenas (a bitmap “asset” file
194 mapping regions of the recorded image to specific experimental arenas) and an auto-
195 reference process that removes images of actual animals from the immobile background
196 (thus allowing them to be traced using the actual experiment). The auto-reference stage
197 was programmed into each assay and lasted 10 minutes (locomotion and Y-maze assays) or
198 5 minutes (light-off startle assay) to make sure that within this interval each individual had
199 moved in its arena.

200 Experimental procedures in the units are controlled by a scripting language (Zanscript) –
201 scripts describing the three assays used in our study can be found in the GitHub repository
202 (https://github.com/elmacartney/Dmel_methods).

203 We have used 4 phenotyping units in total. Three of them were used to run Y-maze assays
204 (40 minutes), whereas the fourth one was used to run the locomotion (30 minutes) and
205 startle (10 minutes) assays concurrently to the Y-maze assays on another set of flies. Such
206 arrangement of units means that different batches of flies were subjected to subsequent
207 tests in different orders (Y-maze was the first performed assay in some flies while
208 locomotion and startle assays were first in others). Any potential variation introduced by
209 differing orderings of the assays is captured by appropriate grouping variables that can be
210 introduced as random effects in statistical analyses.

211

212 *Data analysis*

213 The light-off startle and locomotion data were used without additional data collation, i.e.,
214 the respective responses (see above for details) were used in downstream procedures. Since
215 the files produced by Zantiks units have a particular format (a header section with technical
216 details, followed by the actual data that is formatted according to the script run on the
217 machine, then a footer), data processing involved parsing each file to extract the most
218 relevant information (e.g., experimental unit ID, assay ID, run date, formatted data matched
219 with the numbers of experimental arenas). Parsing steps used in each assay type are
220 presented in the GitHub repository (https://github.com/elmacartney/Dmel_methods;
221 please see the *R* directory for a detailed RMarkdown document). Pre-processing of the
222 startle assay outputs were based on an earlier study applying the same test [25]. The details of
223 file parsing may differ if the Zanscript responsible for performing the experiment is modified
224 or additional data is being generated: in such cases, it is recommended to test the parsing
225 section of the code to ensure correct columns are extracted from the raw data files.

226 The Y-maze behavioural test outputs required more sophisticated processing. Our protocol
227 is based on a modified analysis from Cleal *et al.*[29]. In brief, the analysis extracted all maze
228 arm switches observed during the assay (i.e., transitions between two maze arms, termed
229 'zone changes'). The switches were then classified as left (L)- or right (R)-turns, and
230 sequences of consecutive L/R-turns extracted from a sliding window moving along the
231 recorded sequence of Y-maze locations were assembled into trigrams. We then calculated
232 the proportion of alternating (LRL or RLR) vs. sequential (LLL, RRR) vs. partial (RRL, LLR)
233 movements in the total count of all possible maze explorations (which also include returning
234 to the same zone).

235 Resulting response variables were analysed using linear mixed-effects models with a
236 gaussian response (continuous variables, e.g., the distance travelled by an individual within
237 an arena) or generalised mixed-effects models with a Poisson error distribution (count
238 variables, e.g., the number of alteration sequences in a Y-maze). Mixed models were fitted
239 using the *lme* function in *nlme* package[30] in R (R Core Team, 2021). Each model contained
240 fixed effects of sex and date. Experimental batch ID was included as a random effect. For the
241 models analysing locomotion activity, fly movement was quantified in five repeated
242 measurement bins of 10 minutes. In these models we also included the random effect of
243 individual fly ID to accommodate for this source of dependence. Continuous variables were
244 log-transformed wherever needed (based on the visual inspection of model residuals) and
245 zero-centred. To test for potential effects of fly fatigue we have also generated an additional
246 variable indicating whether a given assay was performed as the first or second assay on a
247 given batch of flies (in all cases, the flies were either tested in the Y-maze, and then for
248 locomotion and startle response, or in the reversed order; locomotion always preceded the
249 startle assay). Finally, we also ran a multivariate model including the locomotor activity,

250 alternation and repetition indices, and handedness (i.e., asymmetry in choosing left versus
251 right runs in the Y-maze) to estimate cross-trait correlations and demonstrate that
252 usefulness of our protocol in identifying potential behavioural syndromes. The model was fit
253 in MCMCglmm using an uninformative inverse-Wishart prior for variance and flat normal
254 priors for means.

255 **Results**

256 *Locomotion tracking*

257 Individual flies exhibited ample variation in their locomotor activity, as measured by the
258 total distance travelled by each individual within the 10-minute bins (Fig. 2A, Fig. S1).

259 Locomotion behaviour was repeatable across five consecutive assay bins: a linear mixed
260 model looking at locomotor activity and partitioning variation in this trait into the between-
261 and within-individual variance indicated the existence of substantial repeatability (ICC =
262 0.36, LRT $\chi^2 = 412.6$, $df = 1$, $p < 0.001$). Experimental block explained 17% of the overall
263 variance. When allowed to differ between the sexes via a sex-specific heterogenous
264 (co)variance structure (assuming cross-sex correlations of zero), inter-individual variance
265 tended to be larger in males than in females (Table 1; LRT $\chi^2 = 16.6$, $df = 1$, $p < 0.001$).
266 Residual variance was also heteroscedastic between the sexes (Table 1; LRT $\chi^2 = 34.3$, $df = 1$,
267 $p < 0.001$.) Due to our pilots reasonably small sample size (sample size can easily be
268 increased by extending the number of batches; also see General remarks), this result should
269 however be treated with caution as the model had problems reaching convergence.

270 Males and females differed in their overall locomotor activity, with males being notably
271 more mobile, in addition to exhibiting greater variance in activity (Table 1 and Fig. 2B). Flies
272 not moving at all may be indicative of fly exhaustion linked to prolonged handling. We
273 tested for this by checking whether the fraction of flies not moving at all in the locomotion
274 assay (distance = 0) differed between batches of flies tested as first or second. Fraction of
275 zero-distance individuals differed between sequence positions (0.6% for flies tested first for
276 locomotion, 8.9% for flies tested second; $\chi^2 = 76.8$, p < 0.001, $df = 1$). However, when

277 included in the linear model for distance covered, sequence positions did not generate any

278 differences (estimate: -0.03, t = -0.22, df = 350, p = 0.81).

279

280

281 *Y-maze*

282 Pure repetition (e.g., LLL, RRR) and alternation (e.g., LRL, RLR) behaviours were the most
283 prevalent (Fig. 3A), with alternations having the highest frequency of all trigram types.

284 Both proportions of repetitions and alternations were highly variable at the between-
285 individual level (Fig. S2). They also exhibited a high fraction of zeroes (i.e., individuals that
286 did not exhibit repetition or alternation behaviours; see General remarks for possible
287 discussion of this issue, and the Discussion for details; Fig. S2).

288 Average locomotor activity tended to be positively associated with the overall proportion of
289 repetition trigrams (Fig. S3 A), a pattern that was absent for the proportion of alterations
290 (Fig. S3 B).

291 When expressed as the absolute number of alternation behaviours (rather than as a
292 proportion), the occurrence of alternation patterns was strongly positively related to
293 average locomotor activity of individual flies, and the strength of this relationship varied
294 between sexes (Table 2, Fig. 3B).

295 Sexual dimorphism was visible in all types of Y-maze behaviour (Fig. S4). Interestingly, the
296 sexes tended to differ in the extent of biased handedness (i.e., asymmetry in choosing left
297 versus right runs in the Y-maze; Fig. S4 C). Batches of flies tested in the Y-maze as second in
298 the sequence tended to be more explorative as they produced significantly more alteration
299 trigrams (GLM with Poisson error: 1.72, $Z = 19.42$, $p < 0.001$) and repetition patterns (0.53, Z
300 $= 7.34$, $p < 0.001$).

301

302 *Light-off startle*

303 Patterns in startle response habituation were not as clear as in the other two measured
304 variables. Most importantly, in over 60% of cases the flies did not exhibit any measurable
305 movement response to the light-off stimulus (i.e., the movement score was zero across all
306 three light-off pulses). Subsequent analyses included only individuals that moved
307 significantly in any of the three pulses (i.e., that had non-zero movement distance when
308 summed across the three light-off pulses).

309 The overall pattern indicated that many individuals, according to expectation, decreased
310 their movement response during the second exposure to the light-of stimulus. However,
311 this "habituation" trend was rarely maintained during the third exposure (Fig. 4, see also Fig.
312 S5).

313 Formal analysis using the post-stimulus distance data revealed no clear decreasing linear
314 trend among the three consecutive stress exposures (Table 3). Males did not differ from
315 females in their average startle response nor did the sexes differ in their habituation slope
316 (Table 3).

317 Although a random slopes analysis might not be numerically stable and robust with the
318 sample size used in this pilot, a simple extension of a mixed model used to analyse
319 habituation to include random slopes (i.e., slopes of the habituation response randomly
320 varying between measured individuals) indicated the existence of non-negligible individual
321 variation (LRT comparing the intercept-only and random slopes models: $\chi^2 = 10.36$, $df = 1$, p
322 $= 0.005$).

323 **Discussion**

324 Our pilot study and its results indicate that certain types of behaviour in small model
325 organisms such as *Drosophila melanogaster* are amenable to automatic phenotyping and
326 can become a promising, new avenue in high-throughput analysis of animal behaviours.
327 Here we show that coupling accurate movement tracking with custom-made scripting
328 language creates a powerful system capable of measuring and recording several types of
329 behaviours that can be relevant in evolutionary and ecological contexts. Our study also
330 indicates that repeated measurements on the same individuals are logistically and
331 technically possible (but also see General remarks for comments on how to improve this
332 aspect of a *Drosophila melanogaster* high-throughput phenotyping study). To date, similar
333 analyses were done on considerably smaller numbers of individuals, mostly due to technical
334 considerations. Increasing numbers of assayed individuals brings such assays to a
335 completely new level, where evolutionary questions centred on between-individual
336 variability and its components can be addressed. Note that the number of individuals used
337 here were designed as a pilot study and it is possible to increase the sample size even
338 further in more formal, hypothesis-driven studies (see General remarks).

339 Of all three applied assays, the general locomotor activity assay was the most
340 straightforward and generated the best quality data. It also exhibited the highest levels of
341 between-individual differentiation and lowest levels of potentially difficult statistical issues
342 (such as zero-inflation of the response). This can, in part, be due to the simplicity of this
343 assay, but also due to the relatively longer measurement period (i.e., ensuring that at least
344 some non-zero amount of the measured behaviour is observed during the assay). Our
345 results clearly suggest that there are sexual differences both in the average levels of this
346 behaviour, and in the between-individual variability (repeatability). Locomotor activity can
347 be used as a measure of exploratory behaviour[32]. Its deterioration could also be used in

348 toxicity, thermal limits, aging, and other similar assays where mobility and movement
349 patterns of an organism are used as quantitative proxies of an individual's response to
350 experimentally applied stress.

351 The Y-maze exploration assay demonstrated that the proposed method also allows for real-
352 time processing of the behavioural data, e.g., to generate Y-maze zone data. Such assays
353 could easily be expanded (e.g., by equipping the Y-maze with olfactory and other sensory
354 stimuli), providing high-throughput ways of assaying learning behaviours, aversion
355 responses or effectiveness of knockouts targeted to specific sensory-regulating genes[33,34].
356 Our Y-maze output data were considerably zero-inflated, which could be an intrinsic feature
357 of the data (and hence could be modelled with appropriate zero-inflated
358 Poisson/binomial/beta distributions) or could result from technical considerations (see
359 General remarks). Nevertheless, Y-maze exploration indices tended to correlate with overall
360 levels of movement activity of the flies, suggesting consistency in behavioural patterns
361 measured by different assays.

362 The most challenging assay, the light-off startle analysis, gave the weakest and most
363 ambiguous results. Surprisingly, a large portion of flies in this assay did not react to the
364 stressful light-off stimulus at all, raising questions about the generality of this assay as
365 presented in earlier studies[25]. Part of the lack of uniformity in the data from this assay and
366 previously published accounts could be due to the flies not being entirely synchronised in
367 terms of their age (i.e., while all the eggs used to generate the focal flies were laid within 24
368 hours of each other, flies could have varied in the time it took for them to eclose into
369 adults). Youth and uniform age seem to be some of the more important factors in
370 determining the success of the light-off response habituation assay[24].

371 Because the studied traits were relatively simple, we do not argue that they represent a
372 one-to-one correspondence with complex behaviours occuring in wild contexts. Characters
373 considered in our paper should be regarded as proxies, or isolated components, of more
374 complex behaviours. Although likely not seen in natural settings (e.g., exploration of a small,
375 confined space; movement in a narrow, unbaited triple-arm maze), these traits represent
376 simplified behavioural units that expose variation underlying more complex behaviours.
377 More importantly, all measured traits seem to be weakly to moderately linked together
378 (Table 4, all correlations but one are statistically significant) – which supports their
379 relevance, and emphasises the potential of the proposed protocol in identifying broader
380 behavioural syndromes.

381 *General remarks*

382 Our report clearly emphasizes that high-efficiency phenotyping of labile characters – such as
383 multiple behavioural traits – is possible even in small and delicate organisms such as the
384 fruit fly. Combining automatic tracking in real-time with carefully designed miniature arenas
385 and flexible experiment-scripting language, as applied in our study, provides a new
386 approach to the challenge of large, high-throughput phenotyping of individual small animals
387 such as *Drosophila melanogaster* (commonly used due to their fast lifecycle). Following our
388 tests, we have further refined the methods applied in this study to increase the
389 reproducibility, efficiency, and ease of the performed procedures. Some of the
390 improvements that could be considered include:

391 - **Using inbred or otherwise isogenic lines of flies instead of repeatedly testing**
392 **individual animals.** Repeated tests on individuals (e.g., performing the same assay
393 across multiple time points in their lifespan or performing sequential tests on
394 specific individuals) can be problematic as flies are delicate and easy to damage

395 during transfer between different arenas. Such handling also requires repeated
396 exposure to cold or CO_2 to induce a temporary coma. Our data indicate that some
397 fatigue effects may be present (flies tested for locomotion at the end showed an
398 excess of individuals not moving at all), but effects are not as obvious as one might
399 expect (flies tested last in Y-mazes showed more exploratory behaviour). However,
400 similarly replicated data can be obtained by replacing individual (i.e., genetically
401 distinct) flies with multiple individuals coming from inbred lines, e.g., the Drosophila
402 Genetic Reference Panel [35,36]. In such a case, individual flies from one line can be
403 deemed unique genotypes (e.g., 'individuals' if maintained under identical
404 conditions) and so can be tested only once in a given type of behavioural assay.
405 - **Considerable time can be saved by housing flies in individual vials divided into**
406 **complete batches to be assayed on a given testing occasion.** Doing so greatly
407 reduces the time needed to isolate the required pool of tested flies and removes the
408 need of anaesthetising a large population of flies to subsample it for a specific
409 number of individuals. Such a strategy also makes it easy to age-synchronise all
410 individuals, which may be key to reducing nuisance variability and improving the
411 reproducibility of certain tests (e.g., the light-off startle assay [24]).
412 - **Refraining from using CO^2 to anesthetise flies can reduce the time needed for the**
413 **animals to regain full activity**[37]. Instead, using an ice bath to stun the flies and cool
414 down all working surfaces (e.g., the arenas/wells into which the flies are loaded)
415 provides enough immobilisation without compromising flies' activity or neuronal
416 performance in the longer term.
417 - **Flies should be assayed outside of their mid-day low-activity period** which may be
418 difficult when performing large numbers of assays on many flies. To avoid this, we

419 propose ensuring that batches are run early morning and late afternoon or by
420 batching the flies into several smaller cohorts maintained at 12:12 photoperiods
421 shifted in relation to each other by a few hours. By shifting the photoperiod, flies are
422 always tested in their 'morning' even when the assays are run over midday. We
423 suspect that increased zero-inflation (i.e., excess of immobile flies) observed in
424 several variables measured in this study could be due to assaying some of the flies
425 after their morning activity peak. Targeting the (relative) morning hours in all
426 measured flies and standardising the measurement time window with respect to the
427 fly circadian cycle should provide improvements in their overall activity levels and
428 eliminate unnecessary sources of trait variability.

429 - **Using mutants altered in target traits linked to the assayed behaviours** could be an
430 effective validation method that would enable stricter identification of behaviour
431 components playing role in our protocols.

432 In summary, we demonstrate that the fruit fly – a fast growing and fast reproducing model
433 organism – provides a study system highly suitable for large, high-throughput phenotyping,
434 which will likely lie at the centre of variation-focused behavioural biology. We provide a
435 suite of simple assays that can easily be conducted in most *Drosophila melanogaster*
436 laboratories with minimum adjustments to standard procedures and protocols. Proposed
437 methods are cost-effective, logistically flexible, and can be modularised to improve their
438 robustness. We believe that further development of such techniques will soon become one
439 of the key elements of evolutionary and behavioural biology, where targeting variance-
440 related questions requires sample sizes and replication levels not achievable with
441 conventional phenotyping approaches.

442 **Acknowledgements**

443 We would like to thank members of the Inter-Disciplinary Ecology and Evolution Lab (I-DEEL)
444 and our colleagues in the Evolution and Ecology Research Centre at the University of New
445 South Wales, Sydney for their support. We would also like to thank Willy Lin and Zina
446 Hamoudi (University of Sydney) for their helpful advice regarding rearing the flies used in
447 this experiment.

448 **Funding**

449 This research was funded by the Australian Research Council DECRA Fellowship
450 (DE180100202) awarded to SD and the ARC Discovery grant (DP200100367) awarded to SN.

451 **Declaration of interest statement**

452 The authors have no competing interests to declare.

453 **Data availability statement**

454 All data and code can be found on Github at
455 https://github.com/elmacartney/Dmel_methods

456 **References**

457 1. Manning A, Dawkins MS. An introduction to animal behaviour. Cambridge University Press;
458 1998.

459 2. Shettleworth SJ. Cognition, evolution, and behavior. Oxford university press; 2009.

460 3. Nakagawa S, Schielzeth H. Repeatability for Gaussian and non-Gaussian data: a practical
461 guide for biologists. Biological Reviews 2010; 85:935–56.

462 4. Dingemanse NJ, Wolf M. Recent models for adaptive personality differences: a review.
463 Philosophical Transactions of the Royal Society B: Biological Sciences 2010; 365:3947–58.

464 5. Conrad JL, Weinersmith KL, Brodin T, Saltz JB, Sih A. Behavioural syndromes in fishes: a
465 review with implications for ecology and fisheries management. J Fish Biol 2011; 78:395–435.

466 6. Dingemanse NJ, Dochtermann NA, Nakagawa S. Defining behavioural syndromes and the role
467 of ‘syndrome deviation’in understanding their evolution. Behavioral Ecology and Sociobiology
468 2012; 66:1543–8.

469 7. Sih A, Bell AM, Johnson JC, Ziemba RE. Behavioral syndromes: an integrative overview. The
470 quarterly review of biology 2004; 79:241–77.

471 8. Macartney EL, Zeender V, Meena A, de Nardo AN, Bonduriansky R, Lüpold S. Sperm depletion
472 in relation to developmental nutrition and genotype in Drosophila melanogaster. Evolution (N
473 Y) 2021; :1–12.

474 9. Macartney EL, Bonduriansky R, Crean AJ. Frequent mating reduces male mating rate but not
475 offspring quality or quantity in a neriid fly. Evolutionary Ecology 2020; 34:915–27.

476 10. Baier A, Wittek B, Brembs B. Drosophila as a new model organism for the neurobiology of
477 aggression? Journal of Experimental Biology 2002; 205:1233–40.

478 11. Miller N, Gerlai R. Quantification of shoaling behaviour in zebrafish (Danio rerio). Behavioural
479 brain research 2007; 184:157–66.

480 12. Koricheva J, Jennions MD, Lau J. Temporal trends in effect sizes: causes, detection and
481 implications. In: Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of Meta-analysis in
482 Ecology and Evolution. Princeton: Princeton University Press; 2013. page 237–54.

483 13. Festing MFW. Randomized block experimental designs can increase the power and
484 reproducibility of laboratory animal experiments. ILAR J 2014; 55:472–6.

485 14. Festing MFW, Altman DG. Guidelines for the design and statistical analysis of experiments
486 using laboratory animals. ILAR J 2002; 43:244–58.

487 15. Cauchoix M, Chow PKY, van Horik JO, Atance CM, Barbeau EJ, Barragan-Jason G, Bize P,
488 Boussard A, Buechel SD, Cabirol A, et al. The repeatability of cognitive performance: a meta-
489 analysis. Philosophical Transactions of the Royal Society B: Biological Sciences 2018;
490 373:20170281.

491 16. Dingemanse NJ, Dochtermann NA. Quantifying individual variation in behaviour: mixed-effect
492 modelling approaches. Journal of Animal Ecology 2013; 82:39–54.

493 17. Balaban-Feld J, Valone TJ. Changes in courtship behaviour following rejection: The influence
494 of female phenotype in Drosophila melanogaster. Ethology 2018; 124:149–54.

495 18. Lupold S, Manier MK, Ala-Honkola O, Belote JM, Pitnick S. Male Drosophila melanogaster
496 adjust ejaculate size based on female mating status, fecundity, and age. Behavioral Ecology
497 2011; 22:184–91.

498 19. Bath E, Morimoto J, Wigby S. The developmental environment modulates mating-induced
499 aggression and fighting success in adult female Drosophila. Functional Ecology 2018;
500 32:2542–52.

501 20. Dukas R. Learning affects mate choice in female fruit flies. Behavioral Ecology 2005; 16:800–
502 4.

503 21. Durisko Z, Dukas R. Effects of Early-Life Experience on Learning Ability in Fruit Flies. Ethology
504 2013; 119:1067–76.

505 22. Maggu K, Kapse S, Ahlawat N, Arun MG, Prasad NG. Finding love: fruit fly males evolving
506 under higher sexual selection are inherently better at finding receptive females. Animal
507 Behaviour 2022; 187:15–33.

508 23. Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P. Automated monitoring and analysis
509 of social behavior in Drosophila. Nat Methods 2009; 6:297–303.

510 24. Fenckova M, Blok LER, Asztalos L, Goodman DP, Cizek P, Singgih EL, Glennon JC, IntHout J,
511 Zweier C, Eichler EE, et al. Habituation learning is a widely affected mechanism in Drosophila
512 models of intellectual disability and autism spectrum disorders. Biol Psychiatry 2019; 86:294–
513 305.

514 25. Allen TA, Budenberg WJ. Replicating Light-Off Startle Responses in Drosophila melanogaster.
515 bioRxiv 2021;

516 26. Simonnet MM, Berthelot-Grosjean M, Grosjean Y. Testing Drosophila olfaction with a Y-maze
517 assay. JoVE (Journal of Visualized Experiments) 2014; :e51241.

518 27. Buchanan SM, Kain JS, de Bivort BL. Neuronal control of locomotor handedness in Drosophila.
519 Proceedings of the National Academy of Sciences 2015; 112:6700–5.

520 28. Akhund-Zade J, Ho S, O'Leary C, de Bivort B. The effect of environmental enrichment on
521 behavioral variability depends on genotype, behavior, and type of enrichment. Journal of
522 Experimental Biology 2019; 222.

523 29. Cleal M, Fontana BD, Ranson DC, McBride SD, Swinny JD, Redhead ES, Parker MO. The Free-
524 movement pattern Y-maze: A cross-species measure of working memory and executive
525 function. Behav Res Methods 2021; 53:536–57.

526 30. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. Linear and nonlinear mixed effects models.
527 R package version 2007; 3:1–89.

528 31. R Core Team. R: A language and environment for statistical computing [Internet]. 2020;
529 Available from: http://www.r-project.org/

530 32. Russell PA. Relationships between exploratory behaviour and fear: a review. British Journal of
531 Psychology 1973; 64:417–33.

532 33. Lainiola M, Procaccini C, Linden A-M. mGluR3 knockout mice show a working memory defect
533 and an enhanced response to MK-801 in the T-and Y-maze cognitive tests. Behavioural brain
534 research 2014; 266:94–103.

535 34. Hasan G. Molecular cloning of an olfactory gene from Drosophila melanogaster. Proceedings
536 of the National Academy of Sciences 1990; 87:9037–41.

537 35. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire
538 MM, Cridland JM, et al. The Drosophila melanogaster genetic reference panel. Nature 2012;
539 482:173–8.

540 36. Mackay TFC, Huang W. Charting the genotype--phenotype map: lessons from the Drosophila
541 melanogaster Genetic Reference Panel. Wiley Interdisciplinary Reviews: Developmental
542 Biology 2018; 7:e289.

543 37. Bartholomew NR, Burdett JM, VandenBrooks JM, Quinlan MC, Call GB. Impaired climbing and
544 flight behaviour in Drosophila melanogaster following carbon dioxide anaesthesia. Sci Rep
545 2015; 5:1–10.

546

547 **Appendices**

548 **Fig. S1.** Variation in individual locomotor activity in all assayed flies. Each vertical line
549 represents the maximum and minimum activity score from 5 quantification repeats included
550 in the assay. Dots are outliers (i.e., observations outside of the 1.5*IQR).

551 **Fig. S2**. Distribution the observed counts of repetition and alternation trigrams from the Y-
552 maze assay.

553 **Fig. S3.** Sex-specific relationships between the proportions of repetition and alternation
554 trigrams in all possible trigrams (from the Y-maze assay), and the average locomotor activity
555 of each individual (from the locomotor activity assay).

556 **Fig. S4.** Sex-specific patterns in Y-maze behaviour. The plot shows male and female
557 proportions of repetition (A) and alternation (B) trigrams, as well as sex-specific handedness
558 (C; zero means perfect symmetry and random choice of turn direction, negative values
559 signify right-turn bias).

560 **Fig. S5**. Light-off startle behaviour averaged over all assayed flies. Points are arithmetic
561 means of distance covered at each stimulus exposure (1st, 2nd, 3rd; x-axis), bars represent
562 95% confidence intervals.

563

564 **Tables**

565 **Table 1**. Results from a general linear mixed model looking at patterns driving variation in

566 overall locomotor activity. Fixed effects estimates are provided with SEs and relevant p-

567 values. Random effects estimates are reported as relevant SD. Heteroscedasticity was

568 tested using the multiplicative parametrisation (i.e., female residual SD equals 0.612 x 1

569 while male residual SD equals 0.612 x 1.248 = 0.7638).

Fixed effects

Term	Estimate	SE	df	t	p-value
Intercept	-0.355	0.280	1424	-1.265	0.206
Sex (Males)	0.471	0.077	351	6.092	<0.001
Date (Day 2)	0.328	0.395	2	0.830	0.494

Random effects

Term	SD
Individual ID (Males)	0.506
Individual ID (Females)	0.713
Residual	0.612
Residual heteroscedasticity	
Females	1.000
Males	1.248

570

571

572 **Table 2.** Generalised linear model analysis (Poisson error distribution) of the total

573 alternations' count in relation to sex and average individual locomotor activity. All effects

574 are reported with their SE, Z statistics and p-values.

	Estimate	**SE**	**Z**	**p-value**
Intercept	0.294	0.183	1.604	0.109
Sex (Males)	0.793	0.199	3.978	<0.001
Average locomotor activity	0.006	0.001	5.928	<0.001
Sex * Locomotor Activity	-0.005	0.001	-4.783	<0.001

575

576 **Table 3.** General linear mixed model analysis of variation in "habituation" slopes. Fixed

577 effects' estimates are reported with relevant SE, t-statistics, and p-values. Random effects

578 are provided as standard deviations.

Fixed effects

Term	*Estimate*	*SE*	*df*	*t*	*p-value*
Intercept	3.308	1.028	216	3.218	0.002
Sex (Male)	0.890	1.534	104	0.581	0.563
Startle stimulus sequence no.	0.314	0.482	216	0.651	0.516
Sex * sequence no.	-0.811	0.727	216	-1.116	0.266

Random effects

Term	*Estimate*
Individual ID (Intercept)	5.281
Individual ID (habituation slope)	2.579
Intercept-Slope correlation	-0.939
Residual	3.881

579

580 **Table 4**. Correlations (with 95% CI intervals) for a sample of 4 measured traits,

581 demonstrating the potential of identifying behavioural syndromes.

582

	Locomotory activity	Proportion of alternations	Proportion of repetitions	Handedness
Locomotory activity	1	-	-	-
Proportion of alternations	-0.07 [-0.13; -0.01]	1	-	-
Proportion of repetitions	0.12 [0.07; 0.19]	-0.04 [-0.10; 0.03]	1	-
Handedness	-0.14 [-0.21; -0.09]	0.10 [0.04; 0.17]	-0.18 [-0.23; -0.11]	1

583

584

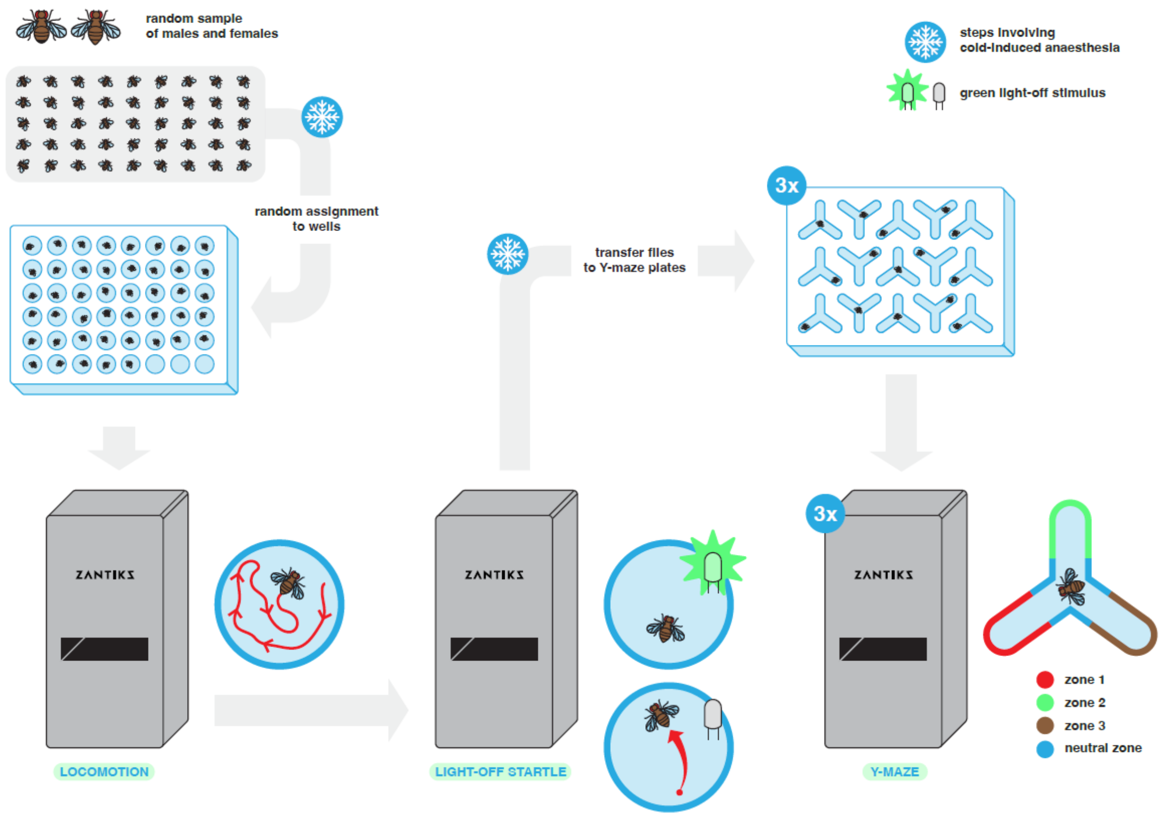
585

586

587

588

589 **Figures**

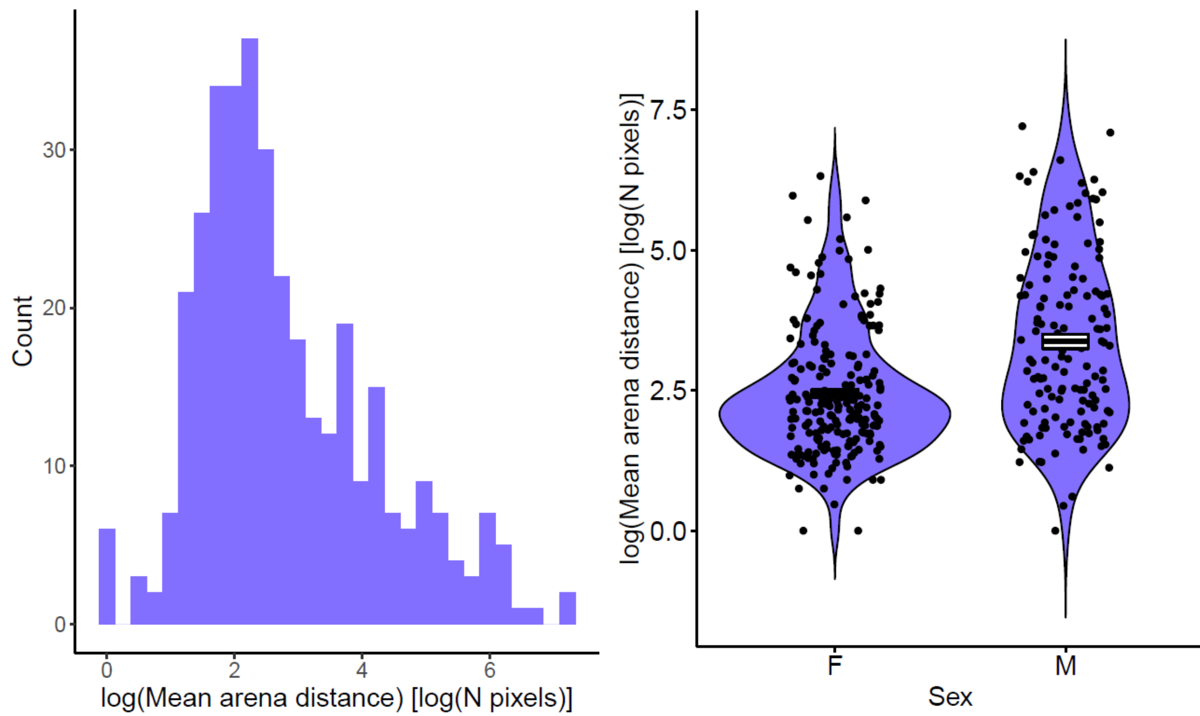


590

591 **Fig. 1.** Overview of the experimental setup demonstrating the steps undertaken with each batch of

592 45 flies.

593

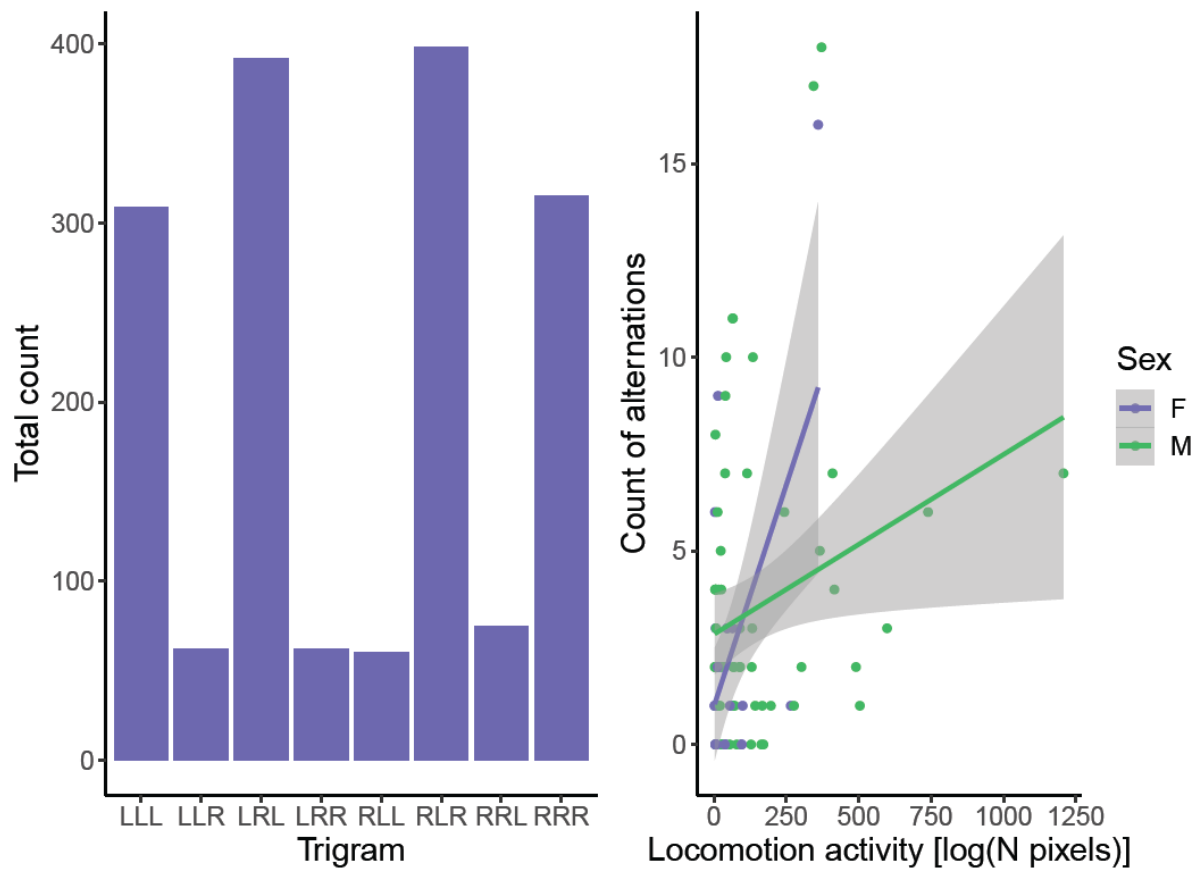


594 **Fig. 2.** Variability in the average locomotion activity (mean of five measurement bins) of *Drosophila*

595 *melanogaster*. The figure shows the histogram of activity for all flies (A) and sex-specific patterns in

596 average locomotor activity (B). In both cases the response was log-transformed.

597



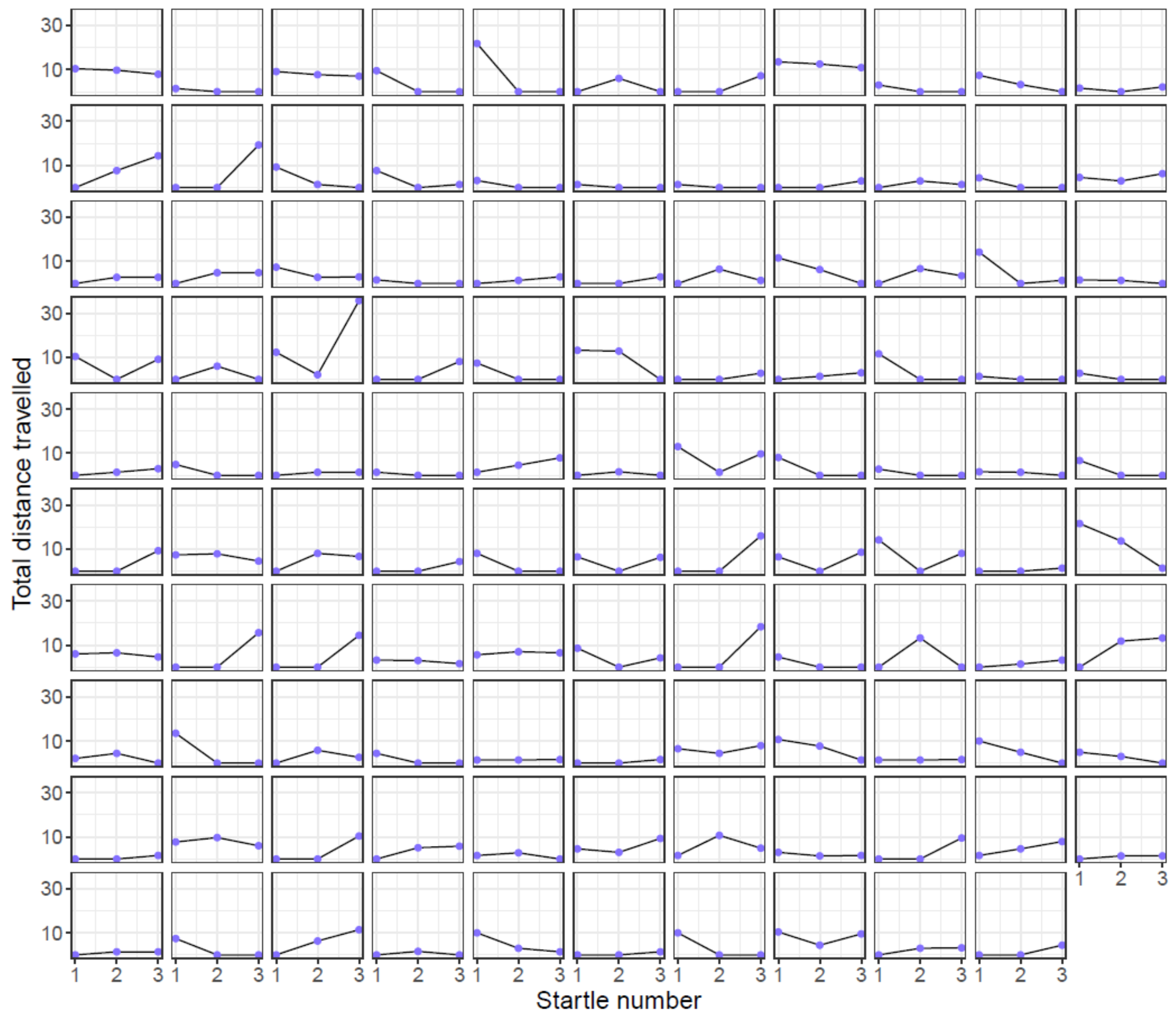
598

599 **Fig. 3.** The distribution of all possible types of trigrams in the Y-maze exploration assay (A) and the

600 relationship between the count of alternation trigrams (LRL and RLR) and the average locomotor

601 activity of each fly (B).

602



603 **Fig. 4.** Habituation to light-off startle response in a sample of flies that exhibited non-zero locomotor
604 activity in at least one of the three measurement timesteps. Horizontal axis of each subplot specifies
605 the three subsequent measurement events; vertical axis presents total distance travelled in the time
606 interval used to register the light-off startle response (measured in number of pixels travelled).