Anwer, H, Mason, D, Zajitschek, S, Noble, D, Hesselson, D, Morris, M, Lagisz, M and Nakagawa, S

An efficient new assay for measuring zebrafish anxiety: tall tanks that better characterize between-individual differences

An efficient new assay for measuring zebrafish anxiety: tall tanks that better characterize between-individual differences

Hamza Anwer[1,2], Dominic Mason[1,2], Susanne Zajitschek[1,2,3], Daniel Noble[4], Daniel Hesselson[2,5], Margaret Morris[6], Malgorzata Lagisz[1,2], Shinichi Nakagawa[1,2]

[1] Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW, Australia

[2] Diabetes and Metabolism Division, Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW, Australia

[3] Liverpool John Moores University, School of Biological and Environmental Sciences, Liverpool L3 3 AF, United Kingdom

[4] Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, Australia

[5] Centenary Institute and Faculty of Medicine and Health, University of Sydney, Sydney, Australia

Abstract

Background: Zebrafish (*Danio rerio*) are increasingly being used to model anxiety. A common behavioral assay employed for assessing anxiety-like behaviors in zebrafish is the “novel tank test”. We hypothesized that using deeper tanks in this test would result in greater between-individual variation in behavioral responses and a more ‘repeatable’ assay.

New methods: After mapping the literature and identifying common behavioral parameters used in analysis, we performed novel tank anxiety tests in both custom-designed ‘tall’ tanks with increased depth and ‘short’ trapezoidal tanks. We compared the repeatability of the behavioral parameters between tall and short tanks and also investigated sex differences.

Results: Overall, regardless of tank depth, almost all behavioral parameters associated with anxiety in zebrafish were significantly repeatable (R = 0.24 to 0.60). Importantly, our tall tanks better captured between-individual differences, resulting in higher repeatability estimates (average repeatability tall tanks: R = 0.46; average repeatability short tanks: R = 0.36) and clearer sex differences.

Conclusions: Our assay using tall tanks has advantages over tests based on short tanks which underestimate repeatability. We argue that use of deeper tanks will improve the reliability of behavioral data across studies using novel tank tests for zebrafish. Our results also call for increased attention in designing the most appropriate assay in biomedical and behavioral sciences as current methods may lack the sensitivity to detect subtle, yet important, information, such as between-individual variation, an important component in assessing the reliability of behavioral data.

Keywords: Zebrafish, Anxiety, Repeatability, Sex, Depth, Dimension

1. Introduction

It is important to infer an animal's internal state to gain insight into why they make certain decisions (Kennedy et al., 2014). Inference into their internal state also provides information regarding the animal's welfare, care requirements, preferences, and dislikes (Mason & Mench, 1997). However, given our inability to directly communicate with animals, inferring internal state is challenging (Corrales-Carvajal et al., 2016), and studying behaviour remains the best option. A range of behavioral assays have been developed and are widely used as important indicators of internal state, such as anxiety (Brown & Bolivar, 2018). Anxiety is defined as "a psychological, physiological, and behavioral state induced in animals and humans by a threat to well-being or survival, either actual or potential" (Steimer, 2002). In humans, anxiety is characterized by excessive worry, hyperarousal, and debilitating fear, and is prevalent worldwide in many population subgroups (Remes et al., 2016). Anxiety is also associated with a range of other health issues (Culpepper, 2009) and places heavy economic burden on affected individuals (Konnopka & König, 2020). Consequently, the importance of anxiety research using animal models has significantly increased over the last several decades (Harro, 2018).

Animal models are powerful for answering anxiety related questions, and are often grouped into two subclasses (Clement & Chapouthier, 1998). The first subclass involves paradigms which assess an animal's conditioned response to aversive stimuli (Freudenberg et al., 2018). The second subclass includes ethological paradigms, which involve the animal's natural reactions to a novel environment (unconditioned response) (Bourin, 2015). The latter attempts to emulate natural conditions under which anxious states are elicited (Bourin, 2015). Classic ethological tests include the open field test (Kraeuter et al., 2019) and elevated plus maze (Pellow et al., 1985). While rodents (rats and mice) are the most commonly used animals in these tests, other animal models have become popular in recent times (Steimer, 2011).

Zebrafish (*Danio rerio*) are increasingly being used as an animal model for addressing anxiety related questions (Blaser & Rosemberg, 2012). They are inexpensive to maintain, reproduce readily and are easy to experimentally manipulate. These features make zebrafish ideally suited for behavioural work provided high-throughput screening methods are available (Nguyen et al., 2013). In addition, they display homologies to humans in key genetic, physiological and behavioral features of stress regulation (Griffiths et al., 2012; Howe et al., 2013; Stewart et al., 2014). Most relevantly, they possess a complex behavioral repertoire (Kalueff et al., 2013) and can be phenotyped to measure their state of anxiety (Stewart et al., 2012). A standard method used to measure zebrafish anxiety is the "novel tank diving test" (ethological paradigm). This method exploits the zebrafish's natural tendency to dive, freeze and reduce exploration in unfamiliar environments (Egan et al., 2009). Typically, the novel environments (tanks) used in zebrafish experiments have limited depth.

However, there seems little to no research on using tanks that have increased depth , despite the fact that zebrafish are known to prefer greater surface depth (Blaser & Goldsteinholm, 2012). We hypothesize that tanks with increased depth will result in more variation in behavioral responses among individuals, and thus provide a more 'repeatable' assay. Repeatability (*R*), also known as intra-class correlation (ICC), is the proportion of phenotypic variation attributed to between-subject (or between-individual) variation (Nakagawa & Schielzeth, 2010). Repeatability is an important index used to quantify the measurement accuracy or constancy of phenotypes. Research has shown that a wide range of behavioral traits are more consistent than previously thought (Bell et al., 2009). This warrants the inclusion of repeatability as an essential index in assessing the accuracy of behavioral studies (Rudeck et al., 2020). Zebrafish display between-individual variation in anxiety; that is, anxiety is a repeatable trait (Thomson et al., 2020). However, an anxiety assay with low sensitivity could fail to adequately quantify between-individual variation. We hypothesized that increasing tank depth in novel tank diving tests would increase between-individual variation, allowing us to develop a more effective assay.

An effective assay can accurately capture differences both between groups and individuals. Behavior is a labile trait (West-Eberhard, 2003) and although it is generally repeatable, on average this repeatability is low, with much of the behavioral variation occurring within individuals, rather than between individuals (Bell et al., 2009). As such, an assay with low repeatability (or high within-individual variation) masks differences between individuals and consequently between groups (that is, variation between two sets of individuals). For example, time spent in the low zone is one of several behavioral parameters used to assess an anxious state in zebrafish in novel tank tests (Maximino et al., 2010). A less effective assay will represent overall behavior as uniform due to the lack of variation between individuals (i.e., all zebrafish are spending similar times at the bottom of the tank). In contrast, an effective assay will capture variable times between individuals and consequently, between groups. Such an assay increases the ability of researchers to make accurate conclusions, for instance regarding treatment efficacy (Senior et al., 2016). Therefore, assays with higher repeatability are usually better able to distinguish differences between groups through greater capturing of between-individual variation or avoiding within-individual variation which overrides behavioral differences among individuals (cf. Fisher et al., 2018; Rudeck et al., 2020).

Here, we describe development of an efficient, new and repeatable anxiety assay for zebrafish. Our main aims for this study are threefold. First, we mapped the literature regarding novel tank anxiety tests in zebrafish. By doing so, we obtained an overview of the main behavioral parameters used to assess anxiety, as well as other information, such as types and dimensions of tanks used. Second, we performed novel tank anxiety tests in both custom-designed 'tall' tanks with increased depth and 'short' trapezoidal tanks. Thus, we examined differences in behavioral parameter measurements (as identified in our survey) between these two types of tanks. Third, we compared the repeatability of the behavioral parameters between tall and short tanks. In addition, we investigated sex differences, as they are ubiquitous and there has been repeated calls for inclusion of sex as an important biological variable in experiments (Jenkins, 2011; Nakagawa et al., 2007).

2. Materials and methods

2.1 Anxiety survey

We performed a systematic review/survey of the academic literature using the online database *Scopus* in May 2020. We used the following search string:

*TITLE-ABS-KEY ("zebrafish" OR "danio rerio" OR "zebra fish" OR "D*rerio") AND ("anxiety-like behaviour*" OR "anxiety-related behaviour*" OR "anxiety test" OR "anxiety assay" OR "tank test" OR "novel tank test" OR "diving test" OR "novel tank" OR "novel tank diving test" OR "video tracking" OR "novel environment" OR "novel tank dive test") AND NOT (bovine OR sheep OR pig* OR drosophila OR cattle OR bull OR vitro OR cow) AND NOT TITLE (women OR men OR patient* OR human* OR child*) AND (LIMIT-TO (DOCTYPE , "ar"))*

Our search in Scopus yielded 336 results. We screened titles and abstracts of downloaded bibliometric records using Rayyan QCRI (Ouzzani et al., 2016). We randomly selected the first 50 experimental studies (Table S8) that met our inclusion criteria. To be included, studies had to be empirical work using laboratory zebrafish in a novel tank test to measure anxiety-like behavior. We then coded experiment-level information from the included studies, such as study focus (e.g. behavioral, medical), treatment (e.g. drugs), and tank type (e.g. rectangular, trapezoidal). We extracted numbers pertaining to tank capacity, tank dimensions, duration of assay and sample sizes, and coded zebrafish behavioral parameters used to assess an anxiety state (available with R code as supplemental files; see section 2.6 below for link). Following extraction, we tallied behavioral parameters and selected seven behaviour measurements (for details, see Results).

2.2 Zebrafish husbandry

Mixed Wildtype (WT) zebrafish stock were raised and maintained in a Tecniplast Zebtec System at 28°C under a 12-h light:12-h dark cycle at the Garvan Institute of Medical Research, Sydney, Australia. Adult zebrafish were housed in 3.5L tanks (max 24 fish per tank in accordance with established Garvan Biological Testing Facility Guidelines GLZ02), and larval zebrafish until 1 month of age were housed in 1.1L tanks (max 40 fish per tank). These housing procedures were also established to reduce impact of dense conditions on growth (Hazlerigg et al., 2012). All tanks received recirculating water (pH 7 – 8 and conductivity 1000 µs) (Aleström et al., 2019). Zebrafish were fed a standard facility diet of Paramecium twice daily, up until 10 – 12 dpf, at which point they were weaned onto live Artemia (twice a day) and dried fish food (once a day). At 60 days post-fertilization (dpf), zebrafish were anesthetized in tricaine solution (4.2 ml of 0.4% in 100 ml of system water) and marked with Visible Implant Elastomer tags (VIE, Northwest Marine Technologies, Inc.; Shaw Island, Washington, United States) for individual identification. We used 9 colored tags: red, brown, purple, black, white, yellow, orange, pink, green; and ‘blank’ (no marking). We injected fish once on either side of the dorsal fin (Hohn & Petrie-Hanson, 2013), unless they were designated blanks. Zebrafish were marked in early November 2019. We used a total of 160 WT zebrafish (n = 79 males, n = 81 females). All the procedures involved in this experiment were approved by the Garvan Animal Ethics Committee (approval: ARA 18_18).

2.3 Testing apparatus

We employed two different tank types (see Figure 1): trapezoidal tanks (width 11 cm, height 17.5 cm, length at top 28 cm, Figure 1B) and custom-designed tall tanks with increased depth (width 7 cm, height 152 cm, length 10.5 cm, Figure 1A). Each tank had a standardized mark displaying the water level at 3.4 L capacity.

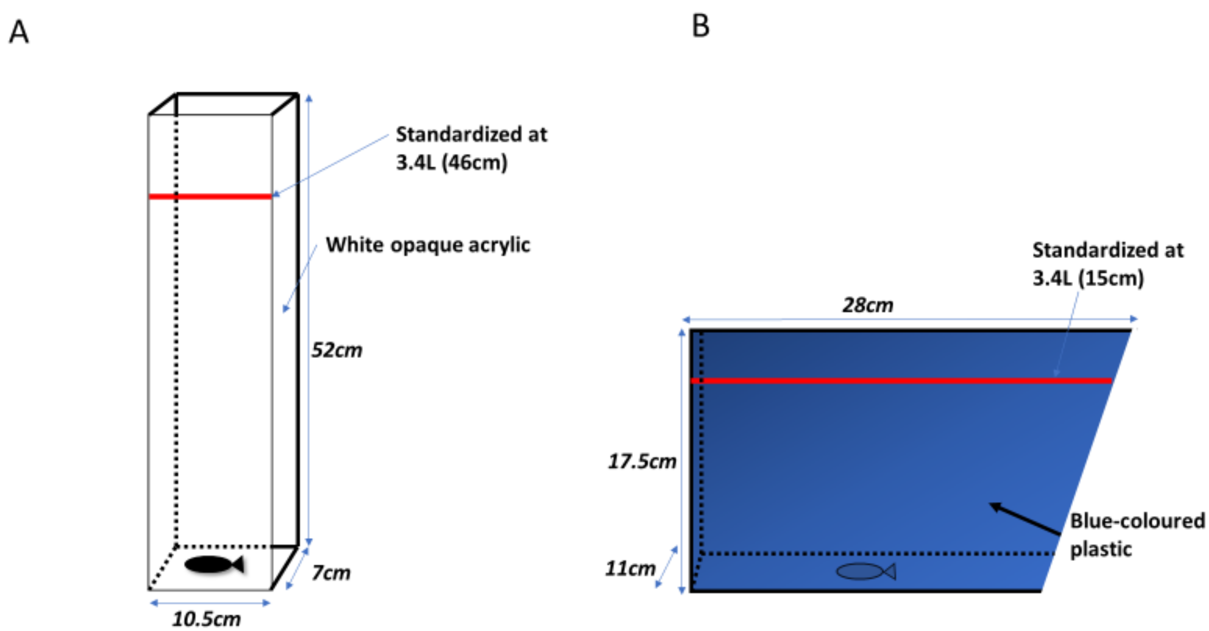

Figure 1 **Characteristics of tanks used in anxiety assays.** A) Our custom-designed tall tank was composed of white opaque acrylic on all sides except the front. The water depth was equated to 46 cm after standardizing the volume of water at 3.4 L; B) Trapezoidal short tanks were composed of blue-coloured transparent plastic.Water depth was equated to 15 cm when the volume was standardized at 3.4 L (hence tanks did not differ in volume of water held).

2.4 Experimental setup

When using tall tanks, we set up 6 tanks to run 6 fish per trial. All 6 tanks were set side-by-side and facing the camera (Figure 2A). White Corflute® sheets were used to block all sides of the arenas except the front portion where the camera was placed; this ensured that fish were not disturbed during trials. When utilizing trapezoidal tanks, we set up 8 tanks to run 8 fish per trial. The setup for the trapezoidal tanks required the use of 2 cameras (4 tanks per camera). To fit 4 tanks in the frame of one camera, we placed a platform (raised approximately 25cm) behind two tanks to place an additional two tanks on top (Figure 2B). A white Corflute® sheet was also placed between the tanks (to prevent fish seeing each other) and behind (to improve contrast). We used the same setup on the other half of the main platform (a Corflute® sheet was placed between both setups). We labelled tanks appropriately with individual fish mark and tank ID.

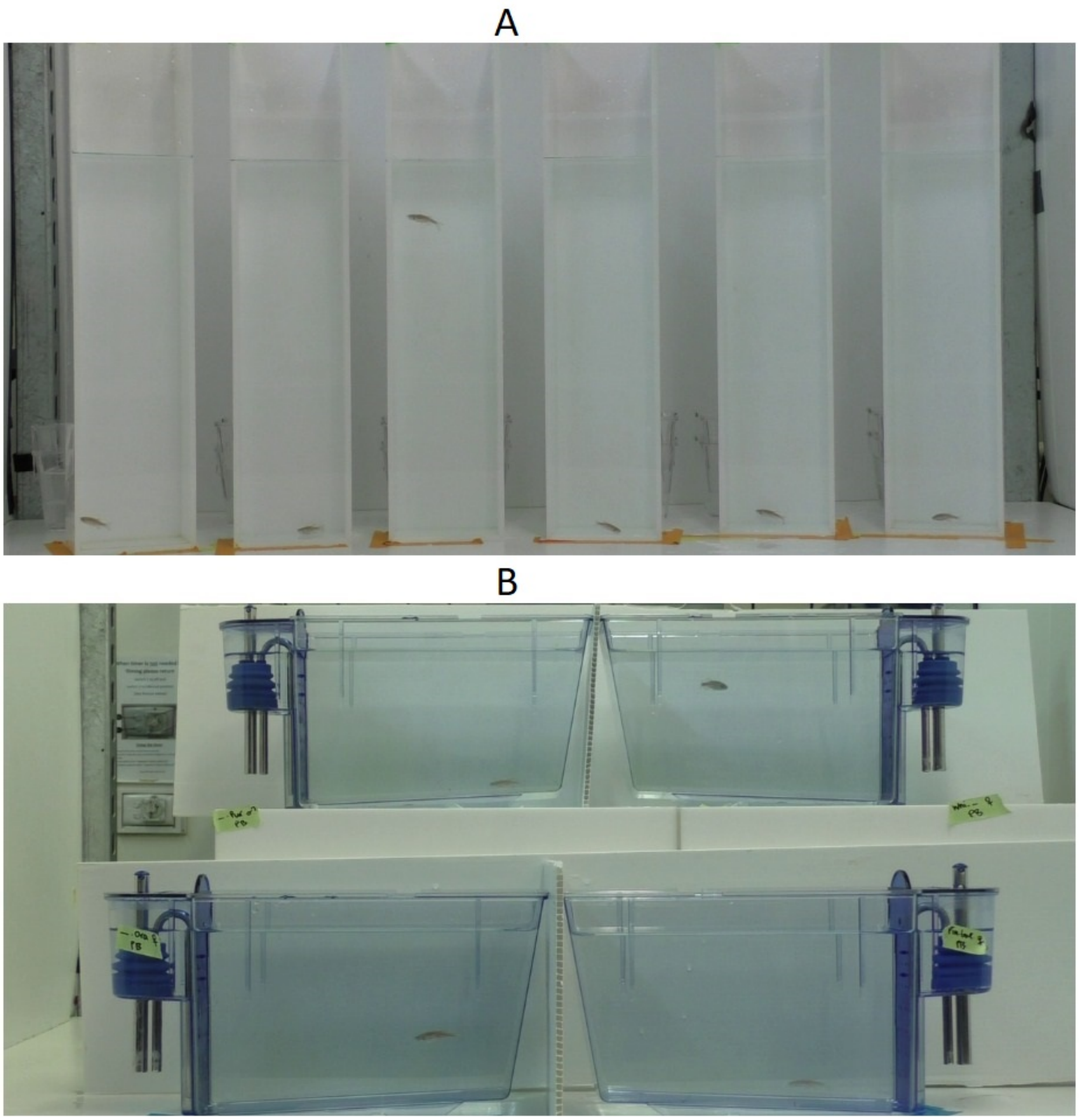

Figure 2 **Setup of tall tanks and short tanks for anxiety assays.** A) Six tall tanks were positioned side-by-side on the main platform. Temporary holding containers were located directly behind each tank for ease of transfer of zebrafish as only one camera was used. No Corflute® was required owing to the opaque acrylic design of the side and back tank walls; B) Four short tanks were positioned in 2 by 2 setup (rows = 2, columns = 2). Those on the top row were placed on a raised custom-made platform. Those on the bottom row were placed directly in front of this platform. This allowed all 4 tanks to be captured in the camera frame. White Corflute® was placed between tanks to prevent fish from seeing each other, as well as behind tanks to improve contrast for video tracking. This same setup was also used on the other half of the main platform. Both halves of the platform were separated by Corflute®. We used 2 cameras to simultaneously capture 8 short tanks at once per trial; we labelled all tanks appropriately with individual fish mark and tank ID.

2.5 Experimental design and procedure

Anxiety assays began in early March 2020. Each individual experienced the anxiety assay in each type of tank twice (i.e. a fish was assayed 4 times in total). For each of the four assay sessions (the sessions were separated by 2 – 3 days), we tested all fish in a single day. We pseudorandomized the order of fish being tested to account for the day of experiments, as well as the time of day. In total, one assay consisted of 20 trials for short tanks (8 fish per trial) and 28 trials for tall tanks (6 fish per trial) (see Supplementary Material for more details). Before each trial, fish were removed from their holding tanks and isolated in separate containers (14 cm × 9 cm × 9 cm; 1.13L) for temporary holding (~5 mins). At the beginning of each assay, fish were transferred from their temporary holding container into their assigned testing tank (tanks 1 – 6 for tall tanks; tanks 1 – 8 for short tanks) and recorded for eight minutes, then removed. This continued until all fish had been assayed for the day. Trials began at 10 am and ended at 4 pm. Water changes occurred every hour to minimize drops in temperature (water was maintained at ~28°C) and the effects of stress hormones from fish already trialed (Pavlidis et al., 2013) (for more details, see the step-by-step protocol in Supplementary Materials).

2.6 Behavioral and statistical analyses

We analyzed all video recordings with the video tracking software Ethovision XT 14.0 (Noldus, Spink and Tegelenbosch, 2001). In Ethovision, we created three digital zones (low, mid and high; see Supplemental Materials Figures S4 and S5) in the tanks for analysis (see Ethovision protocol in Supplementary Materials). Acquisition of data began 40 seconds after the fish had been placed in the testing tank. This was deemed necessary as it took into account the time taken to place all fish in the testing tanks and ensured the lighting and contrast had stablised (changes occured once researchers removed themselves from the frame). We assessed anxiety by analyzing behavioral parameters as decided from our literature survey (see Results).

All statistical analyses were conducted in the R environment (Version 3.4.3) (R Development Team, 2013) with R Studio (Version 1.1.453) (R Studio Team, 2015). To examine mean and variance differences in anxiety-associated behaviour between tall and short tanks, we modelled seven behavioral parameters: 1) time spent in the low zone, 2) time spent in the mid zone, 3) time spent in the high zone; 4) latency to enter the high zone, 5) number of entries into the high zone, 6) total distance travelled, and 7) time spent freezing, with thresholds at 0.25cm/s (start velocity) and 0.10cm/s (stop velocity); see Results on how we chose these behavioral parameters. We used linear mixed models implemented in the function *lme* in the *nlme* package (version 3.1-148) (Pinheiro et al., 2020), which allowed us to model different residual variances. We have used mixed-effects models as they are an overarching framework for ANOVA and t-tests and allowed us to incorporate repeated measurements from the same individuals. This approach has previously been recommended in the field of neuroscience (Aarts et al., 2014; Boisgontier & Cheval, 2016; Shinichi Nakagawa & Hauber, 2011). In addition, mixed models can deal with unequal measurements across individuals when there is missing data (Cnaan et al., 1997). The residual normality of the behavioral measurements was visually checked for all behavioral parameters. We applied transformations to three behavioral measurements to meet the normality assumptions: square-root transformation on time spent in the high zone and entries into the high zone, and ln-transformation on time spent freezing (after adding 1, because of 0 values); these transformed values were used throughout. In all mixed models (seven models; one per behavioral measurement) we used tank type (i.e., our experimental condition) as a fixed factor, as well as water condition (a temporal factor to control for fish being trialed in water that had not yet been changed and therefore exposed to stress hormones from other fish). We used fish ID as a random (clustering) factor. In addition, we also added sex as a fixed factor, as behavioral responses often vary depending on sex (Michelangeli et al., 2016; Schuett et al., 2010) and it is an important biological factor which improves reliability (Tannenbaum et al., 2019). To model different residual variance between tall and short tanks, we specified an *lme* function to do so, but also, we ran the same models assuming a constant variance between the two types of tanks. These two models were compared by likelihood ratio tests using the *anova* function from the R 'stats' package (Version 3.6.2) (R Core Team, 2013) to examine statistical significance for modeling different variances.

Repeatability (R) is formally defined as the proportion of between-group (between-individual) variance out of total variance (Sokal & Rohlf, 2012):

$$R = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$$

where σ_α^2 is the between-group (between-individual) variance and σ_ε^2 is the within-group (within-individual) variance. To calculate repeatability estimates between tall and short tanks, and then between males and females in tall and short tanks, we used *rptR* (Version 0.9.21) (Stoffel et al., 2017), a package based on a mixed-effects model framework using the R package *lme4* (version 20) (Bates et al., 2014). Our repeatability analysis consisted of three steps. First, the overall dataset was divided into tank subsets (i.e. short and tall) to obtain repeatability estimates of each of the seven behavioral measurements with the *rpt* function. We also extracted between-individual and within-individual variance estimates from *rptr* models after performing a *z* transformation on response variables. Second, the dataset was further divided by sex to obtain repeatability estimates of males and females in both tall and short tanks. All estimates were 'adjusted' repeatabilites (Nakagawa & Schielzeth, 2010), and included water condition as a fixed factor and individual fish IDs as a random effect. We obtained standard error and 95% confidence intervals (CIs) using *rptr*, which employs parametric bootstrapping (Faraway, 2016) with all models set to have 10,000 bootstrap samples. Repeatability estimates with confidence intervals not overlapping 0 were considered statistically significant. Third, we calculated contrasts between repeatability estimates. We achieved this by calculating the differences between estimated bootstrap distributions and obtaining quantiles at 2.5% and 97.5% from the difference. Contrasts (subtracting a distribution with a higher mean from that with a lower mean) were deemed significant if the difference distribution did not fall below the 2.5% threshold. All R code and datasets are available at https://github.com/Apex619/Tall_Tanks_Anxiety

3. Results

3. 1 Systematic survey

From 336 studies identified from our literature search, we included 50 for analysis, following our inclusion criteria (Table S8). These studies were published between the years 2008-2020, comprised mainly of behavioral studies (44) with a few medical studies (5) and one toxicology study. Regarding housing tanks used by studies in our sample, 12% housed zebrafish in small tanks (~3-6L), 20% housed zebrafish in large tanks (~100-200L) and 42% housed zebrafish in moderate tanks (~16-50L) (26% of studies did not specify housing tank sizes). Tank types employed were either rectangular in shape (27) or trapezoidal (21), except for two studies (which did not specify the shape). Mean dimensions for rectangular tanks were: height 20.1 cm ± 3.2 SD, width 12.6 cm ± 7.4 SD and length 23.6 cm ± 4.4 SD; and trapezoidal tanks were: height 15.6 cm ± 1.8 SD, width 7.4 cm ± 0.9 SD, length at bottom 22.8 cm ± 0.6 SD and length at top 27.8 cm ± 0.8 SD. Average sample sizes in studies equated to 14 ± 8.9 SD. We identified a total of 16 behavioral parameters from included studies (see Figure 3) and tallied when they were used in included studies. For analysis we used the 6 highly-ranked parameters along with one parameter which was lowly ranked, but we felt was important to include (total of 7 parameters shown in bold; see Figure 3).

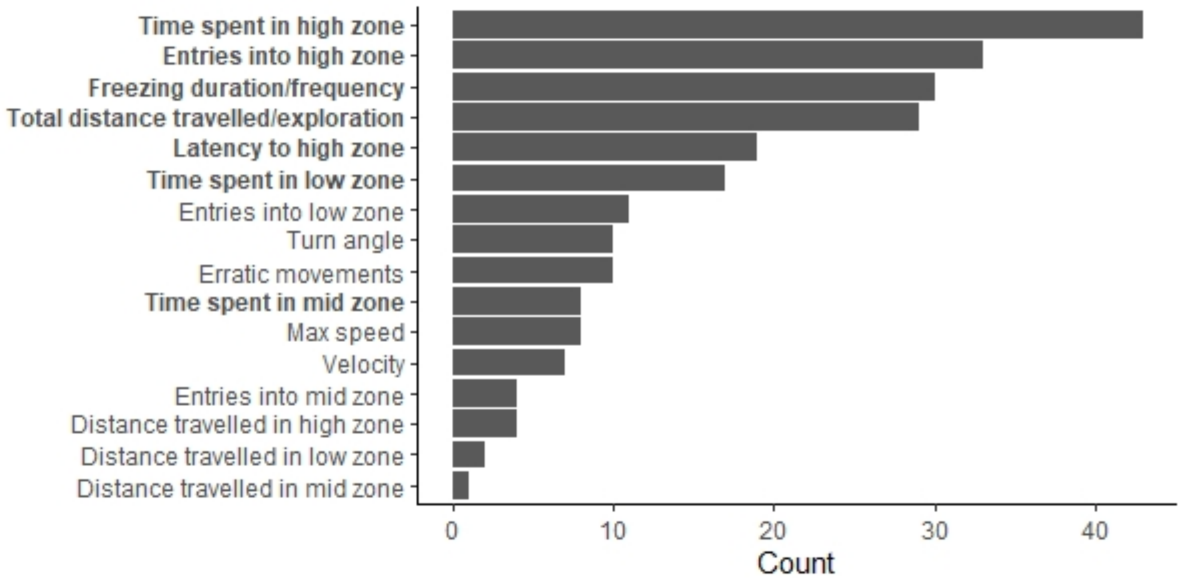

Figure 3 **Results from our systematic survey tallying behavioral parameters used in novel tank test assays from the literature.** From our sample of 50 studies, we identified a total of 16 behavioral parameters used to assess an anxious state in novel tank tests. Of these 16 parameters, we chose 7 (highlighted in bold). The first 6 ranked highest, i.e. were the most frequently used. "Time spent in mid zone" was not amongst the most used parameters, however, we included it based on our design of splitting the tank into 3 zones (as opposed to 2).

3.2 Behavioral parameter measurements

First, statistically significant differences were observed across all behavioral parameters across tank types (see Figure 4, Table S1). In short tanks, zebrafish travelled more (LMM, *est* = 2,323.573, *df* = 469, $t = 25.99$, $p < 0.001$); had longer bouts of freezing (LMM *est* = 1.597, *df* = 469, $t = 8.10$, $p < 0.001$) and spent more time in the low zone (LMM, *est* = 328.927, *df* = 469, $t = 33.34$, $p < 0.001$). In tall tanks, zebrafish spent more time in the mid zone (LMM, *est* = 75.000, *df* = 469, $t = 15.87$, $p < 0.001$) and high zone (LMM, *est* = 8.505, *df* = 469, $t = 18.14$, $p < 0.001$), displayed a quicker latency to enter the high zone (LMM, *est* = 85.123, *df* = 469, $t = 6.36$, $p < 0.001$) and recorded more entries into the high zone (LMM, *est* = 4.365, *df* = 469, $t = 16.61$, $p < 0.001$). Mean responses between sexes did not significantly differ except for the latency to enter the high zone (see Table S1). Water condition had no significant influence on behavioral parameters except for time spent in the low zone and latency to the high zone (see Table S1). Second, tall tanks generated more overall variation than short tanks for time spent in the low zone (6.71%, $p < 0.001$), mid zone (4.47 %, p 0.007) and high zone (6.24%, $p < 0.0001$) as well as entries into the high zone (5.66%, $p < 0.0001$). Time spent freezing however, was more variable in short tanks (4.24%, p 0.0117). No statistically significant differences in variance were observed between tall and short tanks for total distance travelled and latency to the high zone (Figure 4).

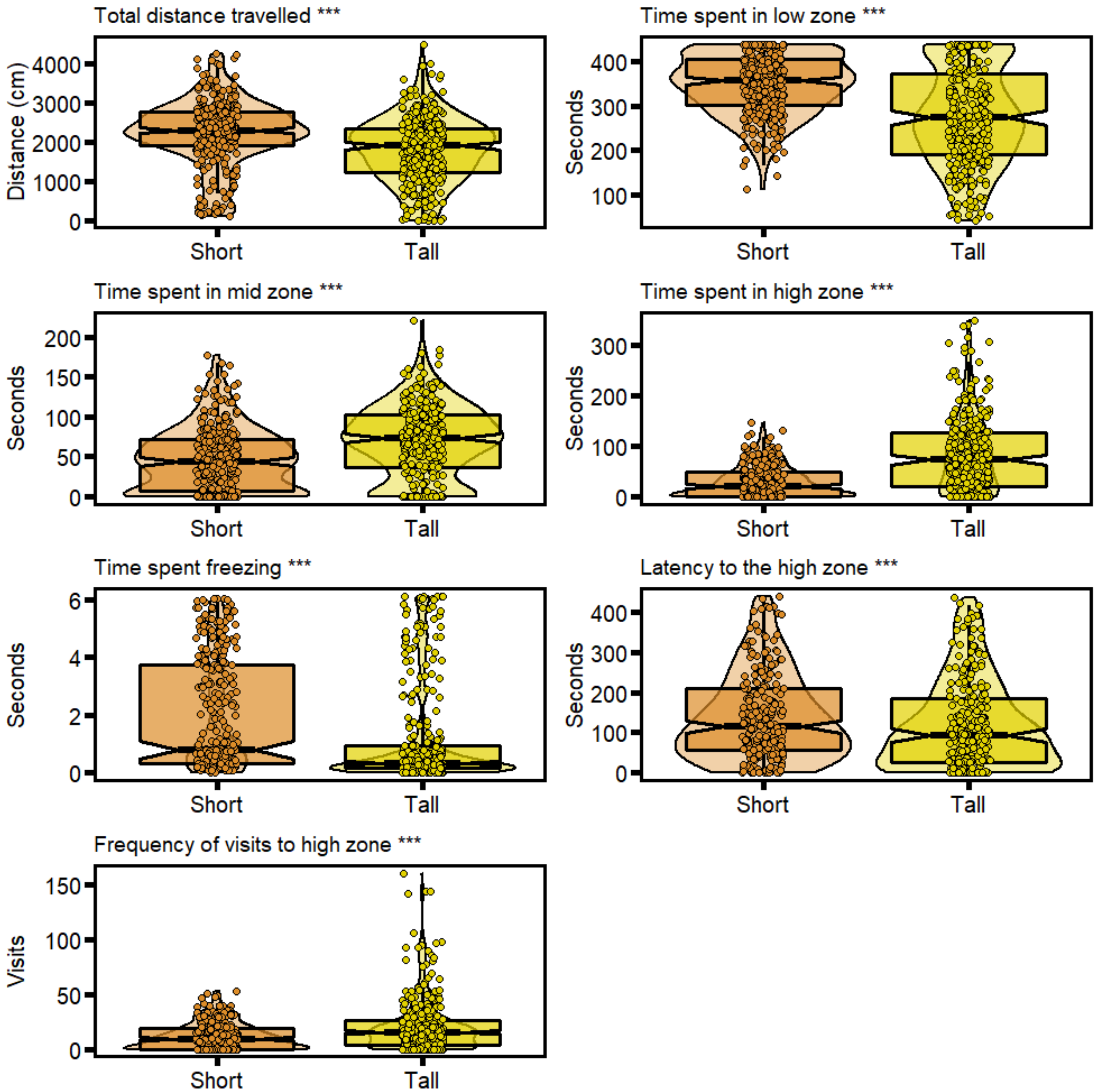

Figure 4 **Distribution of zebrafish behavioral measurements in short and tall tanks.** Each plot displays a combination of: individual data points for males (n = 79) and females (n = 81) from two observations in different tanks (total of 320 observations per plot). Box plots show the median, 95% confidence interval of the median, quantiles and outliers. Violin plots display distribution density. Time spent freezing is transformed using log(x+1) function. Note: $^{***}p<0.01$

3.3 Repeatability analysis

Overall, repeatability estimates were in the expected direction, with tall tanks having higher repeatability than short tanks for 5 out of 7 analysed behavioral parameters (see Figure 5; Table S2): total distance travelled (R = 0.42, 95% CI [0.28 – 0.54]), time spent in the low zone (R = 0.55, 95% CI [0.43 – 0.65]), time spent in the high zone (R = 0.60, 95% CI [0.49 – 0.69]), latency to the high zone (R = 0.49, 95% CI [0.35 – 0.62]) and time spent freezing (R = 0.32, 95% CI [0.18 – 0.45]). However, for only 2 out of these 5 parameters was the difference between tall and short tanks statistically significant: time spent in the low zone (95% CI [0.02 – 0.37]) and latency to the high zone (95% CI [0.13 – 0.58]). Males had higher repeatability estimates than females for all measured behavioural parameters, displaying a clear sex difference (see Figure 6; Table S4). Except for the total distance travelled and time spent freezing, all repeatability estimates in tall tanks were significantly different between males and females.

Short tanks had higher and statistically significant repeatability estimates only for time spent in the mid zone (R = 0.51, 95% CI [0.38 – 0.62]) and entries into the high zone (R = 0.48, 95% CI [0.35 – 0.59]). Results for sex differences were mixed in short tanks (see Figure 7; Table S3). Males had higher repeatability than females for total distance travelled, time spent in the mid zone, and time spent freezing. However, females had higher repeatability than males for time spent in the low zone, time spent in the high zone and entries into the high zone. Repeatability estimates for latency to the high zone in short tanks were statistically non-significant. Unlike in the tall tanks, we only found statistically significant differences between males and females in short tanks for total distance travelled (95% CI [0.38 – 0.70]) and time spent freezing (95% CI [0.03 – 0.56]).

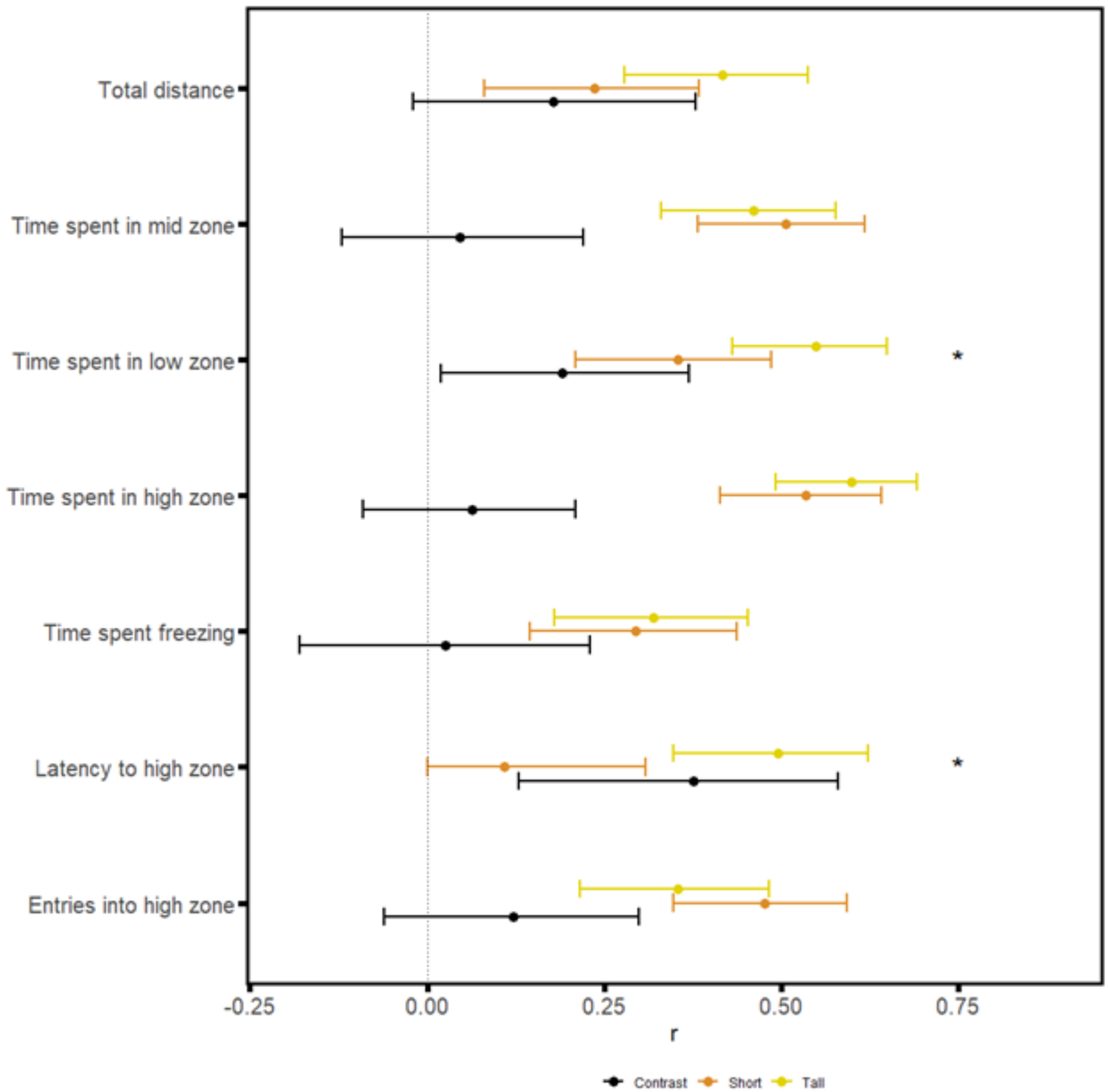

Figure 5 **Forest plot of repeatability estimates for each measured behavioral parameter in tall (yellow) and short tanks (orange), as well as their contrast (in black).** Repeatability estimates are deemed significant if the associated 95% confidence interval does not cross 0. The contrasts are deemed significant (denoted by *) if the associated confidence interval does not cross 0.

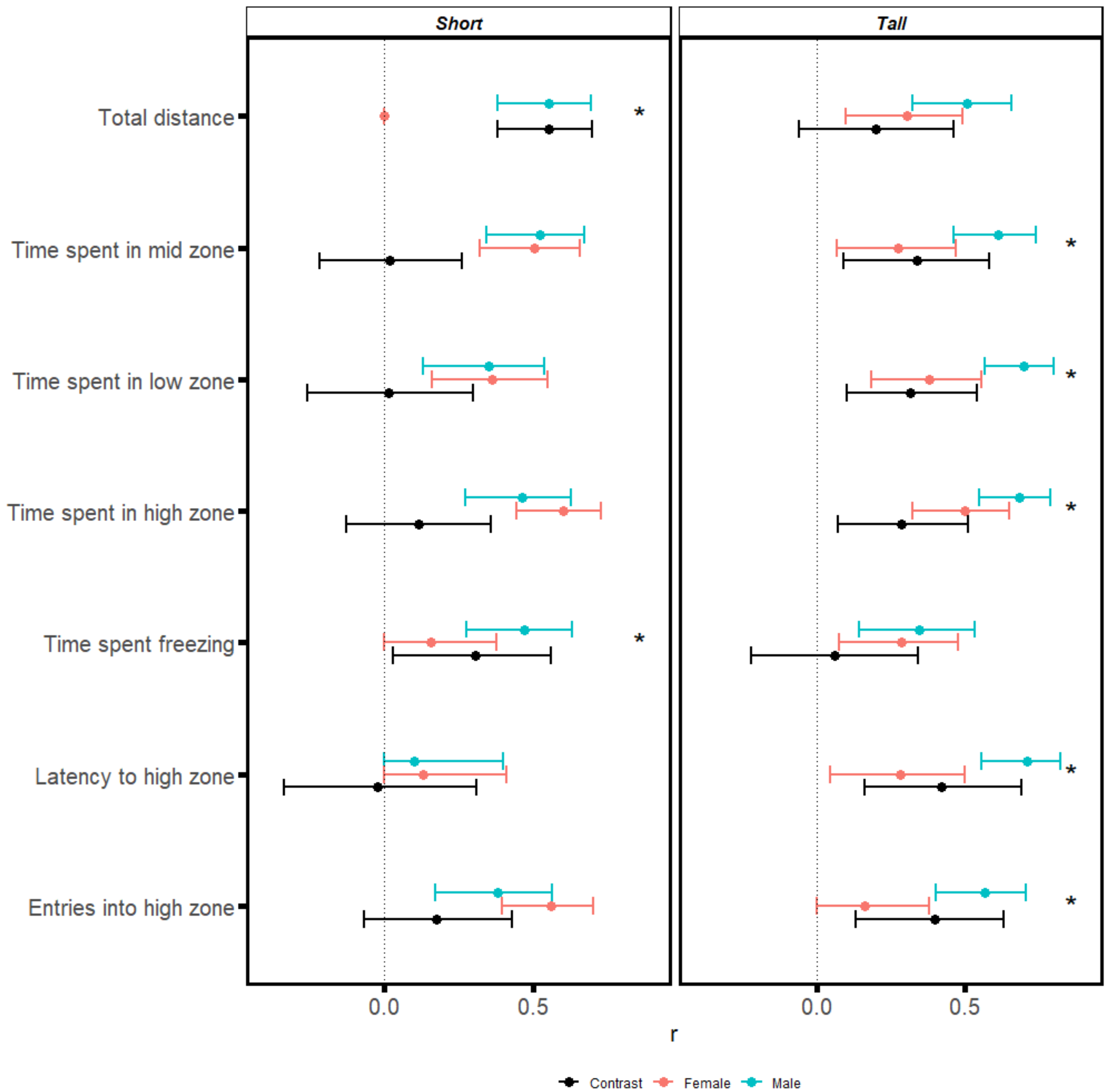

Figure 6 **Forest plot of repeatability estimates for males (blue) and females (red) , as well as the contrast between the sexes (in black), per behavioral parameter in tall and short tanks.** Repeatability estimates are deemed significant if the associated 95% confidence interval does not cross 0. The contrasts are deemed significant (denoted by *) if the associated confidence interval does not cross 0.

4. Discussion

The main goal of this study was to design an efficient anxiety assay that better captures between-individual variation. To do so, we compared the repeatability of behavior in anxiety tank tests between custom-designed tall tanks and short trapezoidal tanks. We addressed three specific aims in this study. First, we mapped a sample of the relevant literature, which confirmed our assumption that studies employ tanks that have a limited depth. Second, we compared anxiety-related behavioral parameters in zebrafish between the two types of tanks, which showed clear behavioral differences. Third, we hypothesized that using the tall tanks would lead to higher repeatability estimates than short tanks. On average, our tall tanks generated more behavioural variation, had higher repeatability estimates and displayed clearer effects between sexes when comparing repeatability estimates. We discuss each of these three points in more detail below.

4.1 Anxiety literature survey

Our survey showed that tanks with depths similar to our tall tanks are not used in novel tank test assays. Although we expected this survey result, it is still somewhat surprising for two reasons. First, when evaluating anxiety, depth is a significant factor in influencing zebrafish behavioral responses (Blaser & Rosemberg, 2012; Córdova et al., 2016; Kysil et al., 2017). Second, anxious zebrafish show a tendency to dive in novel environments (Levin et al., 2007). This diving response indicates a preference to escape the water surface, rather than to simply approach the bottom of a tank (Kysil et al., 2017), emphasizing depth preference. The average heights of tanks used in studies surveyed ranged from 16 to 20 cm (similar to our short tanks which was 17.5cm in height), which may be inadequate in capturing between-individual variation.

4.2 Behavioral response differences

Our analysis revealed that zebrafish in the short tank travelled more and displayed longer bouts of freezing, although both types of tanks had the same volume of water. Total distance travelled may be directly associated with the dimensions of the trapezoidal (short) tank. That is, while shorter in height, the trapezoidal tanks are also much longer in length in comparison to our tall tanks, allowing fish to swim horizontally in the trapezoidal tanks compared to the tall tanks, which limit the fishes' horizontal movements. As such, zebrafish might have adjusted their locomotion to suit this environment (i.e. the tall tank) (Stewart et al., 2012). Furthermore, longer bouts of freezing in short tanks may be the result of a sudden change in social dynamics, as our short testing tanks were the same as those used to house zebrafish in groups (i.e. the novelty may mainly come from social environment disruption rather than the tank itself). Therefore, tanks similar to holding tanks are likely to affect behavioral responses (Bencan et al., 2009).

Overall, we attribute zebrafish behavioral responses to dimensional differences between tall and short tanks. For example, the vertical nature of the tall tank, which had a limited width for horizontal movement, may have driven zebrafish to explore the mid zone and high zone in the tall tank more than in the short tank. As expected, tall tanks also generated more overall behavioral variation than short tanks. This increased variation likely led to enhanced between-individual variation and, consequently, repeatability (see below).

4.3 Repeatability

Overall, we demonstrated that, regardless of tank depths, almost all behavioral parameters associated with anxiety in zebrafish were significantly repeatable in novel tank tests ($R = 0.23$ to 0.60; Figure 5). This result follows suit with a recent study showing significant repeatability in behavioral responses from novel tank tests in zebrafish ($R = 0.35$ to 0.47 for the parameters total distance travelled, time spent in bottom zone, time spent freezing and exploration; Thomson et al., 2020). Indeed, our tall tanks are also better at characterizing between-individual differences by increasing between-individual variation or decreasing within-individual variance (see Figures S1, S2, S3; and Table S2), which results in higher repeatability estimates (tall tanks: $R = 0.30$ to 0.60; short tanks: $R = 0.10$ to 0.53; Figure 5).

Differences in repeatability resulting from the use of tall tanks may have important implications in (bio-)medical science. We argue that seeing too little variation hinders the ability of researchers to make accurate conclusions, for instance regarding treatment efficacy (Senior et al., 2016). Further, identifying and understanding sources of variation is considered necessary to better discern observed responses and better cater treatments at the individual level as opposed to the population level (Braga & Panteghini, 2016; Senn, 2016). More importantly, our new assay, which has higher repeatability, could be more effective in distinguishing effects between control and treatment groups than assays that have lower repeatability (e.g., Mizuno et al., 2020). Essentially, accurately capturing between-individual variation should translate into more accurate capturing of between-group/treatment variation (Fisher et al., 2018). Furthermore, our result highlight the importance of employing methods that ensure behavioral responses are specific to the assumptions of the paradigm being measured, i.e. construct validity (Giuliano et al., 2008; Liu et al., 1982; Maximino et al., 2010). Assays that are usually believed to be appropriate and effective may lack the components needed to detect subtle, yet important, information – including between-individual variation (like what we have shown between conventional short tanks and our custom-designed tall tank).

Our finding also has implications for animal personality studies. Consistent individual differences in behavior (and therefore repeatability) are an essential component of ‘animal personality’ (Dall et al., 2004). Consistent individual differences may represent adaptive behavioral differences within a group (Dall et al., 2004), which, in turn, can influence individual fitness (Dingemanse & Réale, 2005; MacPherson et al., 2017). For example, an animal’s inclination to take risks is associated with the bold-shy behavioral continuum (Sloan Wilson et al., 1994), that is closely related to anxiety (Koolhaas et al., 1999). In novel tank test assays, bold individuals (less anxious) are likely to travel more and traverse to the upper regions of the tank. In our assay, tall tanks captured between-individual variation in behavioral parameters related to total distance travelled and time spent in the low zone better than short tanks (See Figure 5). As previously highlighted, the methodology becomes crucial when attempting to capture between-individual variation.

There seems little emphasis on employing diverse methods to quantify and compare repeatabilities. As such, we call for investing time into comparing and contrasting different assays (e.g., O'Neill et al., 2018) to find the one that is most relevant to the question at hand (note that the most relevant method may not always have the highest repeatability). For example, one way of improving methodology is to assess the ecological relevance of the trait being measured for the species being measured (Roche et al., 2016) (i.e. depth preference in zebrafish, which is better captured by the use of a deeper tank).

We also found significant sex differences in tall tanks, with males displaying more consistent responses than females for all behavioral parameters (tall tanks males: $R = 0.31$ to 0.69; tall tanks females: $R = 0.12$ to 0.49), mimicking results found by Thomson et al. (2020) (males: $R = 0.45$ to 0.58; females: $R = 0.15$ to 0.24). In contrast, results for sexes were mixed in short tanks. Behavioural repeatability was low in females for 3 out of 7 parameters, and there was no clear pattern observed (i.e. one sex being more consistent than the other). However, of the 2 statistically significant results obtained (total distance travelled and time spent freezing), males still displayed higher repeatability than females, a trend also observed in other behavioral studies with different animal models (e.g., Strickland & Frère, 2018; Wexler et al., 2016). Thus, we confirmed the inclusion of sex as an important biological factor to disentangle sources of variation.

4.4 Limitations and future directions

Our improved assay follows the traditional novel tank test. This method relies on zebrafish responding to an unfamiliar environment. However, our assay involved repeated tests in the same tanks making it challenging to maintain tank novelty following the initial assay. This was unavoidable as we aimed to calculate repeatability estimates which required a minimum of 2 measurements. We attempted to ensure that subsequent assays maintained a novelty aspect by 1) having sufficient gaps in between assays (2-3 days) and 2) following a pseudorandomized schedule for the type of tank used (i.e., Day 1 tall tank, Day 2 short tank, Day 3 tall tank, Day 4 short tank). Regardless, we believe the novelty aspect is also caused by a sudden change in social environment (fish are usually housed in groups but then suddenly isolated before and during the assay).

In terms of repeatability, our tall tanks displayed better estimates of repeatability, paving the way for future research to potentially employ our methods. In saying so, our study tested individuals in each tank twice (a total of four assays) over one week (with 2 – 3 days between each assay). However, recent research has highlighted that more tests carried out over an extended period would increase the accuracy of measurements (Thomson et al., 2020). This approach will also address issues associated with observations taken closely together in time, an action which can overestimate repeatability (Mitchell et al., 2020).

Further, our research compared short tanks to custom-designed tall tanks with different dimensions. As such, we did not investigate a ‘truer’ comparison which would have involved comparing short tanks to tanks with identical X-Y dimensions, but with the added feature of increased depth. Our approach was intentional because it provided much greater efficiency given that we were able to film multiple fish at once. In addition, our study would have been confounded due to differences in water volume. Another major strength of our study was our large sample size (79 males and 81 females) in comparison to most studies, enabling us to draw more robust conclusions. However, to ensure all fish were assayed in one day, we employed water changes on an hourly basis rather than a trial-by-trial basis. This would have resulted in some fish being exposed to stress hormones from earlier fish until the water had been changed. To account for this, we included water condition as a factor in our statistical models. While water condition did not significantly influence zebrafish behavioral responses (aside from time spent in the low zone and latency to the high zone), the direction of these responses was biologically consistent with stress. We implore future studies to change water on a trial-by-trial basis or statistically control for water condition to avoid confounds.

In conclusion, our study implemented a custom-designed tall tank to measure zebrafish anxiety in novel tank tests. In doing so, we developed an efficient new assay that captured more between-individual variation, and consequently, repeatability, an important index that improves the reliability of experimental data (Branch, 2019; Hopkins, 2000; Vaz et al., 2013). Also, our tall-tank assay is advantageous in the sense that many studies conducting zebrafish novel tank tests use tanks with limited depth, ranging from ~15-20 cm, whereas our tanks are 46cm deep . Further, our tall-tank assay with increased depth was able to effectively detect sex differences in comparison to our short-tank assay. We highly recommend employing this newly developed assay in anxiety diving tests to improve reliability of behavioral data amongst future studies in (bio-)medical and behavioral sciences.

References

Aarts, E., Verhage, M., Veenvliet, J. V, Dolan, C. V, & van der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. *Nature Neuroscience*, *17*(4), 491–496. https://doi.org/10.1038/nn.3648

Aleström, P., D'Angelo, L., Midtlyng, P. J., Schorderet, D. F., Schulte-Merker, S., Sohm, F., & Warner, S. (2019). Zebrafish: Housing and husbandry recommendations. *Laboratory Animals*, *54*(3), 213–224. https://doi.org/10.1177/0023677219869037

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, *1*(7), 1–23.

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. In *Animal Behaviour* (Vol. 77, Issue 4, pp. 771–783). Academic Press. https://doi.org/10.1016/j.anbehav.2008.12.022

Bencan, Z., Sledge, D., & Levin, E. D. (2009). Buspirone, chlordiazepoxide and diazepam effects in a zebrafish model of anxiety. *Pharmacology Biochemistry and Behavior*, *94*(1), 75–80. https://doi.org/https://doi.org/10.1016/j.pbb.2009.07.009

Blaser, R. E., & Goldsteinholm, K. (2012). Depth preference in zebrafish, Danio rerio: control by surface and substrate cues. *Animal Behaviour*, *83*(4), 953–959.

Blaser, R. E., & Rosemberg, D. B. (2012). Measures of anxiety in zebrafish (Danio rerio): Dissociation of black/white preference and novel tank test. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0036931

Boisgontier, M. P., & Cheval, B. (2016). The anova to mixed model transition. *Neuroscience & Biobehavioral Reviews*, *68*, 1004–1005. https://doi.org/https://doi.org/10.1016/j.neubiorev.2016.05.034

Bourin, M. (2015). Animal models for screening anxiolytic-like drugs: A perspective. *Dialogues in Clinical Neuroscience*, *17*(3), 295–303. www.dialogues-cns.org

Braga, F., & Panteghini, M. (2016). Generation of data on within-subject biological variation in laboratory medicine: an update. *Critical Reviews in Clinical Laboratory Sciences*, *53*(5), 313–325.

Branch, M. N. (2019). The "Reproducibility Crisis:" Might the Methods Used Frequently in Behavior-Analysis Research Help? *Perspectives on Behavior Science*, *42*(1), 77–89. https://doi.org/10.1007/s40614-018-0158-5

Brown, R. E., & Bolivar, S. (2018). The importance of behavioural bioassays in neuroscience. *Journal of Neuroscience Methods*, *300*, 68–76. https://doi.org/10.1016/j.jneumeth.2017.05.022

Clement, Y., & Chapouthier, G. (1998). Biological bases of anxiety. *Neuroscience & Biobehavioral Reviews*, *22*(5), 623–633. https://doi.org/https://doi.org/10.1016/S0149-7634(97)00058-4

Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, *16*(20), 2349–

2380.

Córdova, S. D., dos Santos, T. G., & de Oliveira, D. L. (2016). Water column depth and light intensity modulate the zebrafish preference response in the black/white test. *Neuroscience Letters*, *619*, 131–136. https://doi.org/10.1016/j.neulet.2016.03.008

Corrales-Carvajal, V. M., Faisal, A. A., & Ribeiro, C. (2016). Internal states drive nutrient homeostasis by modulating exploration-exploitation trade-off. *ELife*, *5*(OCTOBER2016). https://doi.org/10.7554/eLife.19920.001

Culpepper, L. (2009). Generalized anxiety disorder and medical illness. *The Journal of Clinical Psychiatry*.

Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004). The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, *7*(8), 734–739.

Dingemanse, N. J., & Réale, D. (2005). Natural Selection and Animal Personality. *Behaviour*, *142*(9/10), 1159–1184. http://www.jstor.org/stable/4536295

Egan, R. J., Bergner, C. L., Hart, P. C., Cachat, J. M., Canavello, P. R., Elegante, M. F., Elkhayat, S. I., Bartels, B. K., Tien, A. K., Tien, D. H., Mohnot, S., Beeson, E., Glasgow, E., Amri, H., Zukowska, Z., & Kalueff, A. V. (2009). Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish. *Behavioural Brain Research*, *205*(1), 38–44. https://doi.org/https://doi.org/10.1016/j.bbr.2009.06.022

Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press.

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115.

Freudenberg, F., O'Leary, A., Aguiar, D. C., & Slattery, D. A. (2018). Challenges with modelling anxiety disorders: a possible hindrance for drug discovery. In *Expert Opinion on Drug Discovery* (Vol. 13, Issue 4, pp. 279–281). Taylor and Francis Ltd. https://doi.org/10.1080/17460441.2018.1418321

Giuliano, S., Director, A., Gambarotta, M., Trasorras, V., & Miragaya, M. (2008). Collection method, season and individual variation on seminal characteristics in the llama (Lama glama). *Animal Reproduction Science*, *104*(2), 359–369. https://doi.org/https://doi.org/10.1016/j.anireprosci.2007.02.016

Griffiths, B., Schoonheim, P. J., Ziv, L., Voelker, L., Baier, H., & Gahtan, E. (2012). A zebrafish model of glucocorticoid resistance shows serotonergic modulation of the stress response. *Frontiers in Behavioral Neuroscience*, *6*(SEPTEMBER). https://doi.org/10.3389/fnbeh.2012.00068

Harro, J. (2018). Animals, anxiety, and anxiety disorders: How to measure anxiety in rodents and why. In *Behavioural Brain Research* (Vol. 352, pp. 81–93). Elsevier B.V. https://doi.org/10.1016/j.bbr.2017.10.016

Hazlerigg, C. R. E., Lorenzen, K., Thorbek, P., Wheeler, J. R., & Tyler, C. R. (2012). Density-

Dependent Processes in the Life History of Fishes: Evidence from Laboratory Populations of Zebrafish Danio rerio. *PLOS ONE*, *7*(5), e37550. https://doi.org/10.1371/journal.pone.0037550

Hohn, C., & Petrie-Hanson, L. (2013). Evaluation of visible implant elastomer tags in zebrafish (Danio rerio). *Biology Open*, *2*(12), 1397–1401.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. In *Sports Medicine* (Vol. 30, Issue 1, pp. 1–15). Adis International Ltd. https://doi.org/10.2165/00007256-200030010-00001

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E. J. J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G.-J. J., White, S., … Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, *496*(7446), 498–503. https://doi.org/10.1038/nature12111

Jenkins, S. H. (2011). Sex differences in repeatability of food-hoarding behaviour of kangaroo rats. *Animal Behaviour*, *81*(6), 1155–1162. https://doi.org/10.1016/j.anbehav.2011.02.021

Kalueff, A. V, Gebhardt, M., Stewart, A. M., Cachat, J. M., Brimmer, M., Chawla, J. S., Craddock, C., Kyzar, E. J., Roth, A., Landsman, S., Gaikwad, S., Robinson, K., Baatrup, E., Tierney, K., Shamchuk, A., Norton, W., Miller, N., Nicolson, T., Braubach, O., … Consortium, Z. N. R. (2013). Towards a comprehensive catalog of zebrafish behavior 1.0 and beyond. *Zebrafish*, *10*(1), 70–86. https://doi.org/10.1089/zeb.2012.0861

Kennedy, A., Asahina, K., Hoopfer, E., Inagaki, H., Jung, Y., Lee, H., Remedios, R., & Anderson, D. J. (2014). Internal states and behavioral decision-making: Toward an integration of emotion and cognition. *Cold Spring Harbor Symposia on Quantitative Biology*, *79*, 199–210. https://doi.org/10.1101/sqb.2014.79.024984

Konnopka, A., & König, H. (2020). Economic Burden of Anxiety Disorders: A Systematic Review and Meta-Analysis. In *PharmacoEconomics* (Vol. 38, Issue 1, pp. 25–37). Adis. https://doi.org/10.1007/s40273-019-00849-7

Koolhaas, J. M., Korte, S. M., De Boer, S. F., Van Der Vegt, B. J., Van Reenen, C. G., Hopster, H., De Jong, I. C., Ruis, M. A. W. W., & Blokhuis, H. J. (1999). Coping styles in animals: current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*, *23*(7), 925–935. https://doi.org/https://doi.org/10.1016/S0149-7634(99)00026-3

Kraeuter, A. K., Guest, P. C., & Sarnyai, Z. (2019). The Open Field Test for Measuring Locomotor Activity and Anxiety-Like Behavior. In *Methods in Molecular Biology* (Vol. 1916, pp. 99–103). Humana Press Inc. https://doi.org/10.1007/978-1-4939-8994-2_9

Kysil, E. V, Meshalkina, D. A., Frick, E. E., Echevarria, D. J., Rosemberg, D. B., Maximino, C., Lima, M. G., Abreu, M. S., Giacomini, A. C., & Barcellos, L. J. G. (2017). Comparative analyses of zebrafish anxiety-like behavior using conflict-based novelty tests. *Zebrafish*, *14*(3), 197–208.

Levin, E. D., Bencan, Z., & Cerutti, D. T. (2007). Anxiolytic effects of nicotine in zebrafish. *Physiology and Behavior*, *90*(1), 54–58. https://doi.org/10.1016/j.physbeh.2006.08.026

Liu, K., Stamler, J., Stamler, R., Cooper, R., Shekelle, R. B., Schoenberger, J. A., Berkson, D. M., Lindberg, H. A., Marquardt, J., Stevens, E., & Tokich, T. (1982). Methodological problems in characterizing an individual's plasma glucose level. *Journal of Chronic Diseases*, *35*(6), 475–485. https://doi.org/https://doi.org/10.1016/0021-9681(82)90062-5

MacPherson, B., Mashayekhi, M., Gras, R., & Scott, R. (2017). Exploring the connection between emergent animal personality and fitness using a novel individual-based model and decision tree approach. *Ecological Informatics*, *40*, 81–92. https://doi.org/https://doi.org/10.1016/j.ecoinf.2017.06.004

Mason, G. J., & Mench, J. (1997). *Using behaviour to assess animal welfare*. Animal Welfare.

Maximino, C., de Brito, T. M., da Silva Batista, A. W., Herculano, A. M., Morato, S., & Gouveia, A. (2010). Measuring anxiety in zebrafish: A critical review. In *Behavioural Brain Research* (Vol. 214, Issue 2, pp. 157–171). Behav Brain Res. https://doi.org/10.1016/j.bbr.2010.05.031

Michelangeli, M., Chapple, D. G., & Wong, B. B. M. (2016). Are behavioural syndromes sex specific? Personality in a widespread lizard species. *Behavioral Ecology and Sociobiology*, *70*(11), 1911–1919. https://doi.org/10.1007/s00265-016-2197-9

Mitchell, D. J., Dujon, A. M., Beckmann, C., & Biro, P. A. (2020). Temporal autocorrelation: a neglected factor in the study of behavioral repeatability and plasticity. *Behavioral Ecology*, *31*(1), 222–231.

Mizuno, Y., McCutcheon, R. A., Brugger, S. P., & Howes, O. D. (2020). Heterogeneity and efficacy of antipsychotic treatment for schizophrenia with or without treatment resistance: a meta-analysis. *Neuropsychopharmacology*, *45*(4), 622–631. https://doi.org/10.1038/s41386-019-0577-3

Nakagawa, S, Gillespie, D. O. S., Hatchwell, B. J., & Burke, T. (2007). Predictable males and unpredictable females: sex difference in repeatability of parental care in a wild bird population. *Journal of Evolutionary Biology*, *20*(5), 1674–1681.

Nakagawa, Shinichi, & Hauber, M. E. (2011). Great challenges with few subjects: Statistical strategies for neuroscientists. In *Neuroscience and Biobehavioral Reviews* (Vol. 35, Issue 3, pp. 462–473). Neurosci Biobehav Rev. https://doi.org/10.1016/j.neubiorev.2010.06.003

Nakagawa, Shinichi, & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, *85*(4), no-no. https://doi.org/10.1111/j.1469-185X.2010.00141.x

Nguyen, M., Yang, E., Neelkantan, N., Mikhaylova, A., Arnold, R., Poudel, M. K., Stewart, A. M., & Kalueff, A. V. (2013). Developing 'integrative' zebrafish models of behavioral and metabolic disorders. *Behavioural Brain Research*, *256*, 172–187. https://doi.org/https://doi.org/10.1016/j.bbr.2013.08.012

O'Neill, S. J., Williamson, J. E., Tosetto, L., & Brown, C. (2018). Effects of acclimatisation on behavioural repeatability in two behaviour assays of the guppy Poecilia reticulata. *Behavioral Ecology and Sociobiology*, *72*(10), 1–11. https://doi.org/10.1007/s00265-018-2582-7

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan---a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Pavlidis, M., Digka, N., Theodoridi, A., Campo, A., Barsakis, K., Skouradakis, G., Samaras, A., & Tsalafouta, A. (2013). Husbandry of zebrafish, danio rerio, and the cortisol stress response. *Zebrafish*, *10*(4), 524–531. https://doi.org/10.1089/zeb.2012.0819

Pellow, S., Chopin, P., File, S. E., & Briley, M. (1985). Validation of open: closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. *Journal of Neuroscience Methods*, *14*(3), 149–167.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2020). *nlme: Linear and Nonlinear Mixed Effects Models* (3.1-148). https://cran.r-project.org/package=nlme

Remes, O., Brayne, C., van der Linde, R., & Lafortune, L. (2016). A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain and Behavior*, *6*(7), e00497. https://doi.org/10.1002/brb3.497

Roche, D. G., Careau, V., & Binning, S. A. (2016). Demystifying animal "personality" (or not): Why individual variation matters to experimental biologists. In *Journal of Experimental Biology* (Vol. 219, Issue 24, pp. 3832–3843). Company of Biologists Ltd. https://doi.org/10.1242/jeb.146712

Rudeck, J., Vogl, S., Banneke, S., Schönfelder, G., & Lewejohann, L. (2020). Repeatability analysis improves the reliability of behavioral data. *PLOS ONE*, *15*(4), e0230900. https://doi.org/10.1371/journal.pone.0230900

Schuett, W., Tregenza, T., & Dall, S. R. X. (2010). Sexual selection and animal personality. *Biological Reviews*, *85*(2), 217–246. https://doi.org/10.1111/j.1469-185X.2009.00101.x

Senior, A. M., Gosby, A. K., Lu, J., Simpson, S. J., & Raubenheimer, D. (2016). Meta-analysis of variance: an illustration comparing the effects of two dietary interventions on variability in weight. *Evolution, Medicine, and Public Health*, *2016*(1), 244–255.

Senn, S. (2016). Mastering variation: variance components and personalised medicine. *Statistics in Medicine*, *35*(7), 966–977.

Sloan Wilson, D., Clark, A. B., Coleman, K., & Dearstyne, T. (1994). Shyness and boldness in humans and other animals. *Trends in Ecology and Evolution*, *9*(11), 442–446. https://doi.org/10.1016/0169-5347(94)90134-1

Sokal, R., & Rohlf, F. (2012). Biometry: the principles and practice of statistics in biological research. 2nd ed. In *Journal of the Royal Statistical Society. Series A (General)* (Vol. 133). https://doi.org/10.2307/2343822

Steimer, T. (2002). The biology of fear-and anxiety-related behaviors. *Dialogues in Clinical Neuroscience*, *4*(3), 231.

Steimer, T. (2011). Animal models of anxiety disorders in rats and mice: some conceptual issues. *Dialogues in Clinical Neuroscience*, *13*(4), 495–506. www.dialogues-cns.org

Stewart, A., Gaikwad, S., Kyzar, E., Green, J., Roth, A., & Kalueff, A. V. (2012). Modeling anxiety using adult zebrafish: A conceptual review. *Neuropharmacology*, *62*(1), 135–143.

https://doi.org/10.1016/j.neuropharm.2011.07.037

Stewart, A. M., Braubach, O., Spitsbergen, J., Gerlai, R., & Kalueff, A. V. (2014). Zebrafish models for translational neuroscience research: from tank to bedside. *Trends in Neurosciences*, *37*(5), 264–278. https://doi.org/10.1016/j.tins.2014.02.011

Stewart, A. M., Gaikwad, S., Kyzar, E., & Kalueff, A. V. (2012). Understanding spatio-temporal strategies of adult zebrafish exploration in the open field test. *Brain Research*, *1451*, 44–52. https://doi.org/10.1016/j.brainres.2012.02.064

Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *8*(11), 1639–1644.

Strickland, K., & Frère, C. H. (2018). Predictable males and unpredictable females: Repeatability of sociability in eastern water dragons. *Behavioral Ecology*, *29*(1), 236–243. https://doi.org/10.1093/beheco/arx148

Tannenbaum, C., Ellis, R. P., Eyssel, F., Zou, J., & Schiebinger, L. (2019). Sex and gender analysis improves science and engineering. *Nature*, *575*(7781), 137–146. https://doi.org/10.1038/s41586-019-1657-6

Team, R. C. (2020). *R: A language and environment for statistical computing*. Vienna, Austria.

Team, Rs. (2020). RStudio: integrated development for R. In *RStudio, Inc., Boston, MA URL http://www. rstudio. com.*

Thomson, H. R., Lamb, S. D., Besson, A. A., & Johnson, S. L. (2020). Long-term repeatability of behaviours in zebrafish (Danio rerio). *Ethology*, *126*(8), 803–811. https://doi.org/10.1111/eth.13038

Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The Case for Using the Repeatability Coefficient When Calculating Test–Retest Reliability. *PLoS ONE*, *8*(9), e73990. https://doi.org/10.1371/journal.pone.0073990

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.

Wexler, Y., Subach, A., Pruitt, J. N., & Scharf, I. (2016). Behavioral repeatability of flour beetles before and after metamorphosis and throughout aging. *Behavioral Ecology and Sociobiology*, *70*(5), 745–753.