Novel methods for discriminating behavioral differences between stickleback individuals and populations in a laboratory shoaling assay

Abigail R. Wark[1,2], Barry J. Wark[2,3], Tessa J. Lageson[1], Catherine L. Peichel[1]

[1]Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle WA 98109-1024, USA, [2]Program in Neurobiology and Behavior, T471 Health Sciences Center, University of Washington, Seattle WA 98195-7270, USA, [3]Present address: Physion Consulting, 11 Worcester Sq #2, Boston MA 02118, USA

Author for correspondence: Catherine L. Peichel

e-mail: cpeichel@fhcrc.org

Phone: (206) 667-1628

Fax: (206) 667-2917

Abstract

Threespine sticklebacks (*Gasterosteus aculeatus*) from different habitats have been observed to differ in shoaling behavior, both in the wild and in laboratory studies. In the present study, we surveyed the shoaling behavior of sticklebacks from a variety of marine, lake, and stream habitats throughout the Pacific Northwest. We tested the shoaling tendencies of 113 wild-caught sticklebacks from thirteen populations using a laboratory assay that was based on other published shoaling assays in sticklebacks. Using traditional behavioral measures for this assay, such as time spent shoaling and mean position in the tank, we were unable to find population differences in shoaling behavior. However, simple plotting techniques revealed differences in spatial distributions during the assay. When we collapsed individual trials into population-level data sets and applied information theoretic measurements, we found significant behavioral differences between populations. For example, entropy estimates confirm that populations display differences in the extent of clustering at various tank positions. Using log-likelihood analysis, we show that these population-level observations reflect consistent differences in individual behavioral patterns that can be difficult to discriminate using standard measures. The analytical techniques we describe may help improve the detection of potential behavioral differences between fish groups in future studies.

Keywords: social behavior, shoaling assay, stickleback, entropy, information theory

Introduction

Social group formation is a common phenomenon among animals and can benefit participants in a number of ways, most notably through facilitating predator avoidance and foraging success (Pitcher and Parrish 1993; Krause and Ruxton 2002). However, these potential benefits depend on ecological and environmental factors such as predation pressure and resource availability. In some cases, social grouping can actually be disadvantageous, either because it is too costly, leading to resource depletion and increased competition, or because it is incompatible with other social behaviors, including courtship, mating, or resource defense (Magurran and Seghers 1991; Krause and Ruxton 2002). Therefore social grouping behavior is expected to differ between animal groups living in different circumstances. In spite of the wealth of literature on the selective forces shaping social aggregation, empirical work investigating variation in group formation among animals is scarce (Krause and Ruxton 2002).

Fish shoals are a popular model system for studying social congregation. Observations of shoaling behavior in guppies and minnows have demonstrated that populations evolving under different predation regimes differ in the strength of social aggregation, with low predation populations forming smaller and less cohesive shoals than high predation populations (Seghers 1974; Magurran 1990; Magurran and Seghers 1991). Lab-raised offspring maintain these behavioral differences, indicating that differences in shoaling behavior between populations are genetically influenced (Seghers 1974; Magurran 1990; Magurran et al. 1995). However, attempts to further elucidate the genetic contributions to shoaling behavior have been inconclusive (Magurran et al. 1992; Wright et al. 2006), potentially hindered by failures to find consistent, repeatable behavioral differences between populations (Parzefall 1993; Wright et al. 2003; Kozak and Boughman 2008). The ability to identify genetic influences on social grouping tendency requires the identification of populations with strong behavioral differences and the use of behavioral measurements that accurately discriminate these differences.

To improve our chances of finding strong differences in social grouping behavior for potential genetic studies, we surveyed the shoaling behavior of thirteen diverse populations of threespine sticklebacks (*Gasterosteus aculeatus*). These small teleost fish have evolved in a variety of isolated aquatic environments that differ in predation regime, food availability, and other ecological characteristics. The behavior, ecology and evolution of this fish has been widely studied, making the stickleback a particularly good model system for studying behavioral adaptations to diverse habitats (Bell and Foster 1994). Previous work on stickleback shoaling behavior has made use of a standard laboratory shoaling assay (Vamosi 2002; Frommen and Bakker 2004; Timmermann et al. 2004; Ward et al. 2004; Wright and Krause 2006; Kozak and Boughman 2008). In this assay, a shoal of fish is isolated at one end of an aquarium tank, and an experimental fish is allowed to swim freely throughout the tank. The shoaling preference of the experimental fish can then be determined by its position in the tank relative to the stimulus shoal. Studies of stickleback shoaling behavior have used this assay to assess the strength of shoaling behavior (tests of shoaling "tendency") (Vamosi 2002; Kozak and Boughman 2008) as well as to identify relevant cues for shoaling (tests of shoaling "preference"), such as body size (Ward et al. 2004) and familiarity (Frommen and Bakker 2004).

In the present study we use this standard laboratory shoaling assay to test whether wild-caught stickleback populations from diverse marine, lake and stream habitats in the Pacific Northwest differ in the strength of their shoaling tendency under standardized laboratory conditions. Using traditional behavioral measures, we fail to find differences in shoaling behavior between the thirteen populations examined. However, using simple plotting techniques and a novel application of information theory concepts to this large behavioral data set, we show that populations do differ in their behavior. We further show that individuals, though variable in their behavior, demonstrate behavioral tendencies that are consistent with their population as a whole. These novel analytic methods provide improved resolution of differences in social grouping behavior among sticklebacks and should be useful for similar studies in other fish.

Materials and methods

Animal collection and care

Adult threespine sticklebacks were collected between May and July 2007 from a variety of locations around the Pacific Northwest. We tested the following populations, listed by habitat type: Marine estuaries: Manchester Clam Bay-MC (n=10), Little Campbell Marine-LM (n=10); Streams: Little Campbell Stream-LS (n=9), Misty Inlet-MI (n=9), Misty Outlet-MO (n=10); Lakes: Beaver Lake-BL (n=7), Hotel Lake-HL (n=10), Misty Lake-ML (n=10), North Lake-NL (n=9), Paxton Benthic-PB (n=10), Paxton Limnetic-PL (n=6), Priest Benthic-RB (n=9), and Priest Limnetic-RL (n=4). These stickleback populations live in habitats that differ in a number of ecological factors, including salinity, water clarity, flow rate, depth, bottom substrate, prey, and predators. Sticklebacks were also collected from Lake Washington (Seattle, WA) for use as a stimulus shoal for behavioral testing.

All fish were caught in unbaited minnow traps, with the exception of North Lake, where fish were collected by hand netting because they could not be caught in traps. Following transport to the lab, individuals from each population were housed together in a single standard aquarium tank under summer lighting conditions (16 hours light, 8 hours dark) at approximately 15.5°C. All tanks contained 3.5 g/L Instant Ocean salt (Instant Ocean, Aquarium Systems, Mentor OH, USA) and 0.4 ml/L $NaHCO_3$, with the exception of the Manchester marine tank, which contained three times more salt.

Fish were caught with permission from the Washington State Department of Fish and Wildlife (07-047) and the British Columbia Ministry of Environment (NA/SU07-31839 and NA07-31713). All animal procedures were approved by the Fred Hutchinson Cancer Research Center Institutional Animal Care and Use Committee (#1575).

Shoaling assay tank

Shoaling trials were conducted between June and July 2007 in the Fred Hutchinson Cancer Research Center's stickleback facility. A behavioral arena (Fig. 1) was constructed in a standard aquarium tank,

with internal measurements of 75 cm long, 46 cm high and 30 cm wide. Two acrylite acrylic dividers (29 cm wide, 40 cm high, 0.5 cm thick) were cemented into the assay with aquarium sealant creating two 10 cm-wide end compartments flanking a 54 cm center arena. To allow water passage between the end compartments and the center arena, each divider had a 0.5 cm hole drilled near each of the four corners. Each hole was 6 cm from the sides, bottom or water line of the tank. A clear cylinder (11 cm diameter) constructed from thin plastic sheeting was placed upright in the center of the tank, creating an isolated acclimation chamber. Fishing line anchored to the rim of the cylinder allowed it to be lifted remotely from behind a curtain without disturbing fish in the trial tank. The assay tank was kept approximately 2/3 full with standard tank water. To maintain a healthy environment in the tank, water was changed every few days and an airstone was placed in the tank between trials. A fluorescent light positioned 22 cm directly above the tank provided uniform lighting in the experimental arena.

Behavioral trials

At the start of a set of trials, twelve individuals were randomly drawn from a laboratory stock of adult wild-caught Lake Washington sticklebacks. Lake Washington sticklebacks were chosen to serve as the shoal stimulus because they are unrelated and unfamiliar to all of the experimental populations. Ten of the twelve Lake Washington sticklebacks were placed in one end compartment and two distracter fish were placed in the other end (Fig. 1). The distracter fish were placed in the tank because pilot experiments performed in our laboratory revealed that when the second compartment contained zero or one fish, all experimental fish, including hypothesized "weak" shoalers, shoaled strongly. Thus, the 10 *vs.* 2 arrangement of shoal fish was designed to reveal variability in shoaling behavior. Any fish that showed signs of being in a reproductive state (gravid belly in females or red throat and blue eyes in males) were not used in the shoal. Blackout curtains were drawn around the tank, and the shoal was allowed to acclimate to the trial tank for 15 minutes. The side of the tank containing the shoal (left or right) was assigned randomly on the first day of testing and was alternated each day thereafter.

Following the shoal acclimation period, the curtains were parted briefly to allow the introduction of an experimental stickleback to the center acclimation chamber. Experimental fish were drawn in arbitrary order from home tanks and introduced directly into the assay tank to minimize the stress of this move. Following a 5-minute acclimation period, during which the experimental fish was able to view the trial tank from within the acclimation chamber, the cylinder was lifted remotely, marking the beginning of the trial. Each trial lasted 15 minutes. At the end of each trial, fish were retrieved, measured and assessed for reproductive status. Each fish was tested once.

We established several *a priori* criteria for discarding potentially erroneous behavioral measures. First, if a fish died within one day of its trial, it was not used for this analysis. One Hotel Lake fish was excluded for this reason. Second, if a fish failed to move from the bottom of the tank within the first five minutes after the start of the trial, the experimenter (watching remotely) ended the trial, removed the fish from the tank, and started the acclimation period for a new trial with a different fish. We excluded 22 fish by this criterion, leaving 113 trials that met our requirements for inclusion in the study.

Video analysis

All trials were recorded using a Sony Handycam digital camcorder (DCR-HC96) positioned 115 cm in front of the assay tank (camera view is depicted in Fig. 1). After all trials had been completed, videos were encoded using QuickTime (Apple Inc., Cupertino CA, USA) and analyzed using custom-built StickleTrack software (Physion Consulting, Boston MA, USA). The position of the experimental fish was recorded once every 3 seconds by measuring the x-y coordinates of the tip of its snout. The z-coordinate (depth from the front to the back of the tank) was not recorded. Horizontal position (x) was recorded as a continuous variable from 0 (at the border of the distracter compartment) to 10 (at the border of the shoal compartment), regardless of whether the shoal was in the left or right compartment. Vertical position (y) was recorded as a continuous variable from 0 at bottom of the tank to 1 at surface of the water.

Statistics

Shoaling behavior in laboratory assays is frequently compared using summary statistics that describe horizontal location in the assay tank, such as mean horizontal position (Vamosi 2002), shoaling time (Wright et al. 2003, 2006), and edge-corrected shoaling time (Timmermann et al. 2004; Kozak and Boughman 2008). In order to compare our work with previous studies of shoaling tendency in sticklebacks (Vamosi 2002; Kozak and Boughman 2008), we calculated these standard shoaling statistics for our data set. Shoaling time and edge-corrected shoaling time measurements require the definition of an area in which a fish is considered to be shoaling with the stimulus. The shoaling time measurement only included time spent near the stimulus shoal compartment (horizontal position 9-10) and did not include time spent near the distracter compartment (horizontal position 0-1). These horizontal positions (0-1 and 9-10) correspond to a physical distance of 0.0 - 5.4 cm from the compartments and represent approximately one body length, as the average length of experimental fish in this study was 5.68 ± 0.79 cm (± standard deviation). Even if they are not shoaling, sticklebacks tend to stay at the edges of the shoaling assay tank rather than in the middle. Therefore some investigators (Timmermann et al. 2004; Kozak and Boughman 2008) correct for this non-shoal-related edge-preference by subtracting time spent near the distracters from total time shoaling. We used time spent within approximately one body length of the distracter compartment (horizontal position 0-1) to correct our shoaling time measurement for non-specific edge-preferences.

All three standard shoaling statistics, as well as median horizontal position, standard deviation of horizontal position, mean vertical position, median vertical position, and standard deviation of vertical position, were calculated for all individuals. Populations were then compared using both Kruskal-Wallis and MANOVA tests, and pairwise comparisons were conducted using Tukey's post-hoc tests. We tested whether the distribution of positions for each individual in the study as well as the population distributions as a whole were Gaussian using the D'Agostino and Pearson Omnibus Test for Normality (D'Agostino and Pearson 1971; D'Agostino and Pearson 1973; Jones et al. 2001). Kolmogorov-Smirnov tests gave similar results. Horizontal (x) and vertical (y) distributions were tested separately.

Information theory analysis

Entropy is a measure of the expected amount of information provided by drawing a sample from a probability distribution (Shannon 1948). The information provided by a single sample, x, drawn from a probability distribution p is given by $-\log p(x)$ and has units of bits when the logarithm is base 2. A single bit of information is the information provided by the result of flipping a fair coin (*i.e.* the probability of heads is 0.5). The entropy (H) of a probability distribution is the average novel information provided by a sample of that distribution:

$$H = \left\langle -\log_2 p(x) \right\rangle_X . \quad (1)$$

The entropy of a distribution is closely related to the variability of that distribution. Intuitively, if a distribution has low variability, the value of a sample from that distribution does not provide much new information—its value could be relatively easily predicted before it was drawn—and the probability distribution has low entropy. Conversely a highly variable distribution makes it more difficult to predict the value of a draw from that distribution so each draw provides significant novel information and the entropy of the distribution is high. The variability of a distribution is commonly measured by the variance—the second central moment—of that distribution. For Gaussian distributions, variance completely describes the variability in the distribution (*i.e.* the value of higher moments such as skewness and kurtosis, the 3rd and 4th moments, are fixed given the mean and variance of a Gaussian distribution). For non-Gaussian distributions, however, variance is not a complete measure of the variability of the distribution. Because we do not know the functional form of the distributions in our data, we would like to use a measure that accounts for variability in all moments of the distribution without making assumptions about the form of the distribution. Entropy is such a measure and is thus a more appropriate and potentially more informative description of the variability of a distribution.

196 To estimate entropy using the binless method (see below), we first need to assume that the probability
197 density, p, of the location of a fish in the assay tank is continuous. The task is then to estimate the
198 differential entropy (H_{diff}) of that distribution,

199 $$H_{diff} = -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx \qquad (2)$$

200 from n independent samples $x_1,\ldots,x_n$ drawn according to $p(x)$. The differential entropy has a fixed, but
201 infinite, offset from the discrete Shannon entropy as defined above. This fixed offset is canceled when we
202 take the difference of two differential entropies. Therefore we ignore the absolute value of the entropy in
203 our analysis and report the difference in entropy between population distributions.

204 Correct estimation of entropy is not trivial, and several methods exist in the literature (Paninski 2003).
205 We used a binless entropy estimator described by Victor (2002). The insight of the binless method is that
206 $p(x_i)$ can be estimated by finding the nearest observed sample to x_i. Intuitively, if the probability
207 density $p(x_i)$ around x_i is high, then we would expect to have observed a second sample near x_i.
208 Conversely, if the probability density around x_i is low, we would expect the nearest observed sample to
209 x_i to be more distant. To estimate entropy using this method, we first change the variable of integration
210 from equation (2) to y, the cumulative probability density $y = \int_{\infty}^{x} p(t)dt$, with $dy = p(x)dx$, giving

211 $$H_{diff} = -\int_0^1 \log_2 p(x)dy\,. \qquad (3)$$

212 Thus, the entropy is expressed as an average log probability where the average is weighted equally with
213 respect to the cumulative probability density.

214 We can estimate the cumulative density from the observed data, where each of N observations
215 contributes $1/N$ to the cumulative density. Equation (3) can then be estimated as

216 $$H_{diff} \approx -\sum_{i=1}^{N} \frac{1}{N}\log_2 p(x_i)\,. \qquad (4)$$

Further, we can estimate $\log_2 p(x_i)$ from the distance to the nearest observation to x_i by estimating the probability $q(\lambda)$ that, after *N*-1 observations, the nearest observation to x_i is at a distance of at least λ. This definition gives $q(\lambda) \approx e^{-\frac{S_r \lambda^r (N-1) p(x_i)}{r}}$ where $S_r = 2\pi^{r/2} / \Gamma\left(\frac{r}{2}+1\right)$, where Γ is the Gamma function, for dimensionality r (r=2 in our analysis). Substituting this result into the definition for $\langle \log_2 \lambda \rangle_\lambda$ gives

$$\langle \log_2 \lambda \rangle \approx \frac{1}{r}\left(-\log_2 \left[\frac{S_r (N-1) p(x_i)}{r} \right] - \frac{\gamma}{\ln(2)} \right) \quad (5)$$

where γ is the Euler-Mascheroni constant (≈ 0.5772156649). Rearranging to solve for $-\log_2 p(x_i)$ and substituting into equation (4) then gives

$$H_{diff} \approx \frac{r}{N} \sum_{i=1}^{N} \log_2(\lambda_i) + \log_2 \left[\frac{S_r (N-1)}{r} \right] + \frac{\gamma}{\ln(2)} \quad (6)$$

where λ_i is the distance from x_i to the nearest observed neighbor. Python code to implement this entropy estimate is provided as electronic supplementary material.

Fish do not jump randomly in space. Therefore the samples of fish position recorded from video are not truly independent. Without taking this correlation into account, the above analysis—and any other statistical analysis that assumes independent samples—will give a biased result. We measured the auto-correlation function of fish position—the correlation coefficient between a given position and the fish's position at a given time delay—for all fish in all populations. We found that this correlation falls off to approximately 1/*e* at 60 seconds in both the *x* and *y* position for all individuals and populations (electronic supplementary material figure S1). As expected, shuffling positions in time eliminates this autocorrelation (data not shown). To produce enough independent samples for our analysis, we took a random sub-sample of the data such that on average we chose only one sample per correlation time (i.e. 60 seconds), giving approximately 15 independent samples per trial for each fish.

Reported entropy estimates are the mean of 500 bootstrapped estimates. The size of the bootstrap sample (n=15) was chosen to produce independent samples of position, as described above. Because the standard error of an estimator is defined as the standard deviation of the distribution of its estimates, the standard error of our entropy estimate is the same as the standard deviation of the bootstrapped estimate.

Estimating population distributions

We estimated the distribution of tracked locations for each population using a Gaussian kernel density estimate (Parzen 1962). A Gaussian kernel density estimate approximates the true distribution with an appropriately normalized sum of Gaussians kernels, each centered on an observed sample. The result can be thought of as a smoothed histogram of observed locations. The kernel estimate has the advantage over a simple histogram estimate of avoiding consideration of how to handle histogram bins with zero observations. For observations $x_1,\ldots,x_n$ the Gaussian kernel density estimate is thus

$$\hat{p}_{KDE}(y) = \frac{1}{n(2\pi)\left|\Sigma_{\text{kernel}}\right|^{1/2}} \sum_{i=1}^{n} e^{-\frac{1}{2}(y-x_i)^{\mathrm{T}}\Sigma_{\text{kernel}}^{-1}(y-x_i)}. \quad (7)$$

Where the samples are more closely spaced, the summed probability density of the Gaussians centered at those points is greater than in areas where samples are widely spaced. The covariance, Σ_{kernel}, of the Gaussian kernel was chosen according to

$$\Sigma_{\text{kernel}} = \zeta^2 \Sigma_{\text{data}} \quad (8)$$

where Σ_{data} is the data sample covariance and ζ is Scott's factor, $n^{-\frac{1}{d+4}}$, for n samples of dimensionality d (Jones et al. 2001). Density heat maps were constructed by evaluating the kernel density estimate on a 100x100 grid of equally spaced locations.

Likelihood analysis

Given an estimate of a population's distribution of locations in the shoaling assay, we wanted to compute the likelihood that an individual fish's tracked locations were drawn from a population's distribution. Given the Gaussian kernel density estimate above, we can calculate the unconditioned probability of a particular individual's tracked location, $\hat{p}_{KDE}(x,y)$, for each tracked location. We estimated the likelihood of a sequence of *n* positions as the product of their independent probabilities, $\prod_{i=1}^{n} \hat{p}_{KDE}(x_i,y_i)$. In results below, we present the related log-likelihood as it is more easily computed using fixed-precision floating point calculations. The log-likelihood of the track is given by $\sum_{i=1}^{n} \log\left(\hat{p}_{KDE}(x_i,y_i)\right)$.

Results

We collected thirteen populations of threespine stickleback from the Pacific Northwest in order to assess inter- and intra- population variation in shoaling behavior using a common laboratory shoaling assay (Fig. 1). We used a standard set of commonly used shoaling measurements, as well as additional measurements, to describe the positional distribution of each individual. We then compared populations using both parametric (MANOVA) and non-parametric (Kruskal-Wallis) tests (Table 1). The populations in our study did not differ in mean horizontal position (Kruskal-Wallis Chi square: χ^2=14.728, *df*=12, *p*=0.257), median horizontal position (χ^2=11.733, *df*=12, *p*=0.467), shoaling time (χ^2=16.595, *df*=12, *p*=0.165), or edge-corrected shoaling time (χ^2=16.154, *df*=12, *p*=0.184). Although we could not reject the hypothesis that the standard deviation of horizontal position in the tank was the same for all populations (Kruskal-Wallis Chi square: χ^2=23.693, *df*=12, *p*=0.022; MANOVA: $F_{(12,100)}$=2.396, *p*=0.009), Tukey's post-hoc tests failed to identify significant pair-wise differences between any populations. Thus, the

stickleback populations we tested in this study did not differ in shoaling behavior according to the standard statistical analyses that are frequently applied to similar data sets (Vamosi 2002; Timmermann et al. 2004; Kozak and Boughman 2008) or according to the additional statistical analyses we used (i.e. median horizontal position, standard deviation of horizontal position).

Interestingly, the thirteen populations we studied did differ in their vertical distribution in the assay tank (mean position: χ^2=48.610, *df*=12, *p*<0.00001; median position: χ^2=51.747, *df*=12, *p*<0.00001; standard deviation: χ^2=23.137, *df*=12, *p*<0.027), though this measure appears to be unrelated to shoaling behavior.

Although standard shoaling measures did not differ among our populations, plotting the raw position data of all fish from each population (Fig. 2a) revealed that the populations exhibit different distributions of positions during the shoaling trials. For example, Paxton Benthics appear to position themselves more uniformly in the tank than Hotel Lake fish. We also observed that the distributions were highly non-Gaussian. This observation was confirmed by failure to meet normality in D'Agostino and Pearson Normality tests. 93% of the animals in the study had horizontal distributions that deviated significantly from normal, while 96.5% of vertical distributions deviated from normal (Table 1). Furthermore, none of the thirteen populations, when tested as summed distributions, met normality in the horizontal dimension, and only one population (Misty Lake) showed a vertical distribution that did not deviate significantly from normality (Table 1).

To quantify observed differences between the populations, we chose an information theoretic measure of variability that is not dependent on a parameterized (*e.g.* Gaussian) model of the distribution. We estimated the differential entropy of the distribution of positions for each population (see Methods). For each population, we performed 500 bootstrap estimates. We found that the populations with low entropy, such as Hotel and Beaver Lakes, are tightly clustered in space, whereas populations with greater entropy, such as Paxton Benthic and Manchester, are less clustered (Fig. 3).

306 A population may have high variability (or entropy) in position during the shoaling assay due to
307 differences between individuals in the population, rather than common population-wide behavioral
308 patterns. Thus, we wanted to test whether the population-level probability distributions (Fig. 2b) and
309 entropy estimates (Fig. 3) are representative of individual behavioral patterns. We computed the
310 likelihood that an individual's positional tracks came from the positional distribution of its own
311 population or a different population. For comparison, we chose the North Lake and Paxton Benthic
312 populations, as the distribution of these populations' positions have non-overlapping entropy estimates
313 (Fig. 3). For each individual from both populations, we computed the likelihood that the fish's sampled
314 trajectory was drawn from the Paxton Benthic or the North Lake population. On average, a fish's
315 trajectory was more likely to have been drawn from its own population (Fig. 4), indicating that the
316 population-level distribution is an appropriate description of individual behavioral patterns.
317
318
319 **Discussion**
320
321 The present study represents a broad survey of diversity in shoaling behavior among stickleback
322 populations. In our study, standard measures of shoaling behavior failed to distinguish any differences
323 among the stickleback populations we studied. However, using a non-traditional set of observational and
324 analytic tools, we do observe differences in the way that different populations behave when tested in this
325 assay. The tools that we have developed using this large data set provide the opportunity for improved
326 resolution of behavioral differences in similar laboratory paradigms. We offer these tools as a resource to
327 the community (see electronic supplementary material) in the hope that they will provide an additional
328 method for comparing behavioral patterns among individuals or populations.
329 Using traditional behavioral measures associated with a common laboratory shoaling assay, we were
330 unable to detect any behavioral differences among wild stickleback populations from a variety of habitats.
331 Two previous investigations of stickleback shoaling tendency assessed differences between Paxton

Benthic and Paxton Limnetic sticklebacks using these same standard shoaling measures. Similar to Kozak and Boughman (2008), we found no difference in tendency to shoal between Paxton Benthic and Paxton Limnetic sticklebacks. However, Vamosi (2002) reported that Paxton Limnetic sticklebacks have mean positions that are closer to the shoal than Paxton Benthic sticklebacks, whose mean positions do not differ from random (the midpoint of the testing tank). Because this original report only used Gaussian statistical summaries to describe individual behavioral distributions, it is impossible to determine whether Paxton Benthic and Limnetic individuals displayed similar behaviors in these independent studies. For example, Paxton Benthic sticklebacks in the Vamosi (2002) study may have displayed a mean position in the center of the tank because they remained in the center of the tank or because they moved throughout the tank, as we observed in the current study (Figs. 3, 4).

Although we failed to find differences in shoaling behavior among stickleback populations using traditional analytical methods, the populations we tested do behave differently in the shoaling tank. Examining the raw positional data of each population (Fig. 2) reveals that these groups differ in the extent to which they position themselves near or away from the shoal, the depth that they maintain in the tank, and the extent to which they are clustered or spread out in the tank. The simple plotting techniques in Fig. 2 highlight two important implications of this study. The first is that a single summary statistic, such as population mean, is insufficient to capture the considerable variation (as well as higher order processes such as skewness or kurtosis) that we observe in this data set. If we hope to be able to compare behavioral data across different studies, the value of presenting raw data alongside any summary analyses cannot be understated. The second observation from the raw data plots in Fig. 2 is that the positional distributions we observe are clearly non-Gaussian, an observation that we confirmed statistically. These data suggest that statistical tests that assume normality are inappropriate for these data.

Having observed different patterns of behavior in the raw data, we set out to characterize and quantify differences among the population distributions. We employed two techniques, entropy estimation and log-likelihood analysis, to assess some of these behavioral differences. Entropy estimates indicate that stickleback populations differ in the extent to which they cluster at consistent positions in the shoaling

assay tank. If we compare low entropy and high entropy populations, we can see that the entropy estimates capture a difference in behavioral pattern that we clearly observe in the positional distribution plots (Fig. 3).

Entropy estimates enabled quantification of differences in behavior among populations, but we could not estimate entropy for individuals due to fewer independent data points. Therefore, we did not know whether these population-level patterns reflected consistent behavioral patterns among individuals within the population or whether they resulted from behavioral differences among individuals within the populations. For example, high entropy estimates could result because all individuals in the population exhibit high scatter or because individuals have low scatter but they position themselves differently from one another. In other words, high entropy could be a symptom of high intra-population variability. To ask whether population-level patterns reflect individual behaviors, we performed a log-likelihood analysis. When we compared populations with different entropy estimates, we see that individuals are more similar to their own population than the alternative population. This result indicates that for populations that differ in entropy, population-level analyses are an accurate reflection of individual behaviors.

Entropy and log-likelihood analyses reveal behavioral differences that are apparent in the distribution plots. However, we do not see any strong ecological or habitat-based explanations for the differences we detect. It is interesting to note that three out of four of the solitary lake populations within the data set (Beaver, North and Hotel Lakes) show the lowest entropy. Paxton Benthic, a population that has been suggested not to shoal (Vamosi 2002), shows the highest entropy score. The study included four pairs of populations that have overlapping distributions but live in ecological divergent habitats (Little Campbell Marine and Stream, Paxton Limnetic and Benthic, Priest Limnetic and Benthic, and Misty Lake and Inlet). Each of these pairs show nearly identical spatial distributions (Fig. 2) and do not differ significantly from one another in entropy (Fig. 3).

To return to the original goal of detecting shoaling differences, can we conclude that populations that differ in entropy also differ in shoaling behavior? Entropy estimates do not necessarily distinguish shoaling from non-shoaling behavior; a population could be a low entropy, non-shoaling population or a

low entropy, shoaling population. Thus, in order to compare relative shoaling behavior, entropy estimates can be used in combination with assessment of positional distributions. In the present analysis, all low entropy populations appear to be shoaling (Figs. 2, 3). Although high entropy populations, such as Paxton Benthic and Manchester, also spend a significant amount of time shoaling, combining the positional histograms with the entropy estimates supports the conclusion that these populations have a weaker shoaling tendency than lower entropy populations.

From this study, we conclude that shoaling behavior, particularly when assessed via the present assay, is not a promising candidate for future genetic analysis in sticklebacks. Nonetheless, the precision and unprecedented size of this survey of shoaling behavior has allowed us to develop more informative techniques for describing and assessing behavioral differences between populations. These techniques may be applicable in a variety of animal behavior paradigms where large positional data sets are collected. For example, studies ranging from the assessment of shoaling, schooling or boldness behavior in the laboratory to much larger spatial and temporal studies that include GPS tracking data in the field, may be amendable to this type of analysis. The information theoretic techniques we describe are publically available (see electronic supplementary material) for use in future studies.

Acknowledgements

We would like to thank Matt Arnegard, Susan Foster, Andrew Hendry, Jean-Sebastien Moore, Dolph Schluter and Mike Shapiro for help collecting sticklebacks, and to Anna Greenwood for advice and support throughout the study. We are also grateful to Adrienne Fairhall and Joe Sisneros for helpful comments on the manuscript. This research was supported by a grant from the National Institutes of Health HG002568 to C.L.P.

References

Bell MA, Foster SA (1994) The evolutionary biology of the threespine stickleback. Oxford University Press, Oxford

D'Agostino RB, Pearson ES (1971) An omnibus test of normality for moderate and large sample size. Biometrika 58:341-348

D'Agostino RB, Pearson ES (1973) Tests for departure from normality. Biometrika 60:613-622

Frommen JG, Bakker TCM (2004) Adult three-spined sticklebacks prefer to shoal with familiar kin. Behaviour 141:1401–1409

Jones E, Oliphant T, Peterson P (2001) SciPy: Open source tools for Python

Kozak GM, Boughman JW (2008) Experience influences shoal member preference in a species pair of sticklebacks. Behav Ecol 19:667-676

Krause J, Ruxton GD (2002) Living in groups. Oxford University Press, Oxford

Magurran AE (1990) The inheritance and development of minnow antipredator behaviour. Anim Behav 39:834–842

Magurran AE, Seghers BH (1991) Variation in schooling and aggression amongst guppy (*Poecilia reticulata*) populations in Trinidad. Behaviour 118:214–234

435 Magurran AE, Seghers BH, Carvalho GR, Shaw PW (1992) Behavioural consequences of an artificial
436 introduction of guppies (*Poecilia reticulata*) in N. Trinidad: evidence for the evolution of anti-
437 predator behaviour in the wild. Proc Biol Sci 248:117-122
438
439 Magurran AE, Seghers BH, Shaw PW, Carvalho GR (1995) The behavioral diversity and evolution of
440 guppy, *Poecilia reticulata*, populations in Trinidad. In: Advances in the study of behavior.
441 Academic Press Limited, London, pp 155–155
442
443 Paninski L (2003) Estimation of entropy and mutual information. Neural Comput 15:1191–1253
444
445 Parzefall J (1993) Behavioural ecology of cave-dwelling fishes. In: Pitcher TJ (ed) Behaviour of teleost
446 fishes. Chapman and Hall, London, pp 573–608
447
448 Parzen E (1962) On estimation of a probability density function and mode. Ann Math Stat 33:1065-1076
449
450 Pitcher TJ, Parrish JK (1993) Functions of shoaling behaviour in teleosts. In: Pitcher TJ (ed) Behaviour of
451 teleost fishes. Chapman & Hall, London, pp 369–439
452
453 Seghers BH (1974) Schooling behavior in the guppy (*Poecilia reticulata*): an evolutionary response to
454 predation. Evolution 28:486–489
455
456 Shannon CE (1948) The mathematical theory of communication. Bell Syst Tech J 27:379–423
457
458 Timmermann M, Schlupp I, Plath M (2004) Shoaling behaviour in a surface-dwelling and a cave-
459 dwelling population of a barb *Garra barreimiae* (Cyprinidae, Teleostei). Acta ethologica 7:59-64
460

Vamosi SM (2002) Predation sharpens the adaptive peaks: survival trade-offs in sympatric sticklebacks. Ann Zool Fenn 39:237-248

Victor JD (2002) Binless strategies for estimation of information from neural data. Physical Review E 66:51903

Ward AJ, Hart PJ, Krause J (2004) Assessment and assortment: how fishes use local and global cues to choose which school to go to. Proc Biol Sci 271:S328-S330

Wright D, Krause J (2006) Repeated measures of shoaling tendency in zebrafish (*Danio rerio*) and other small teleost fishes. Nat Protoc 1:1828-1831

Wright D, Rimmer LB, Pritchard VL, Krause J, Butlin RK (2003) Inter and intra-population variation in shoaling and boldness in the zebrafish (*Danio rerio*). Naturwissenschaften 90:374-377

Wright D, Nakamichi R, Krause J, Butlin RK (2006) QTL analysis of behavioral and morphological differentiation between wild and laboratory zebrafish (*Danio rerio*). Behav Genet 36:271-284

480 **Figure Legends**

481

482 **Fig. 1** Shoaling assay tank. A standard aquarium tank was divided into three compartments: a central
483 testing arena and two end compartments containing the stimulus shoal (n=10) and the distracter fish
484 (n=2). A plastic cylinder that could be lifted remotely provided a temporary acclimation chamber for the
485 experimental fish

486

487 **Fig. 2** Populations differ in their positional distribution in the shoaling assay tank. Each plot displays the
488 complete positional data within the central compartment of the shoaling tank for all individuals within a
489 population. Each plot is presented as though the shoal is located at the right side of the tank. The x-axis
490 ranges from 0 (at left) to 10 (at right); y-axis ranges from 0 (bottom) to 1 (surface). **a** For each
491 population, the position of each individual is plotted as a single dot every three seconds throughout the 10
492 minute trial (300 points per fish). **b** Heat maps constructed using kernel density estimates applied to the
493 data from (**a**) indicate the probability of individuals from each population occupying any given position in
494 the assay tank. Red indicates areas of highest probability and dark blue indicates area of lowest
495 probability

496

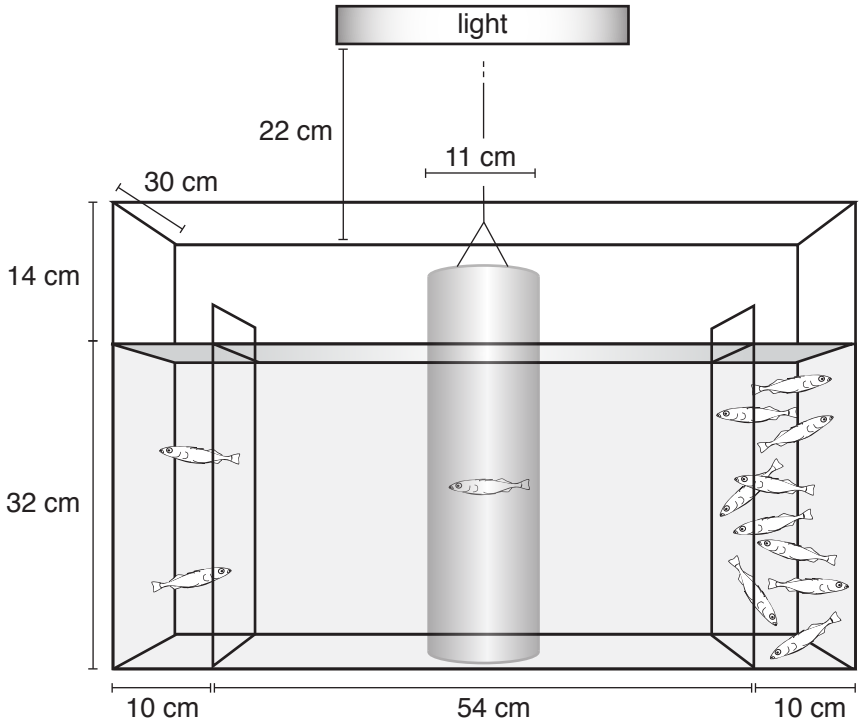
497 **Fig. 3** Difference in entropy between stickleback populations in the shoaling assay. **a** For each
498 population, the estimated entropy from 500 bootstrap estimates is shown. Standard error of the entropy
499 estimate is the standard deviation of bootstrap estimates. Populations are ordered by estimated entropy. **b**
500 Population scatter plots are shown for lowest entropy populations (Beaver Lake and Hotel Lake) and the
501 highest entropy populations (Manchester and Paxton Benthic) (same population graphs as Fig. 2a)

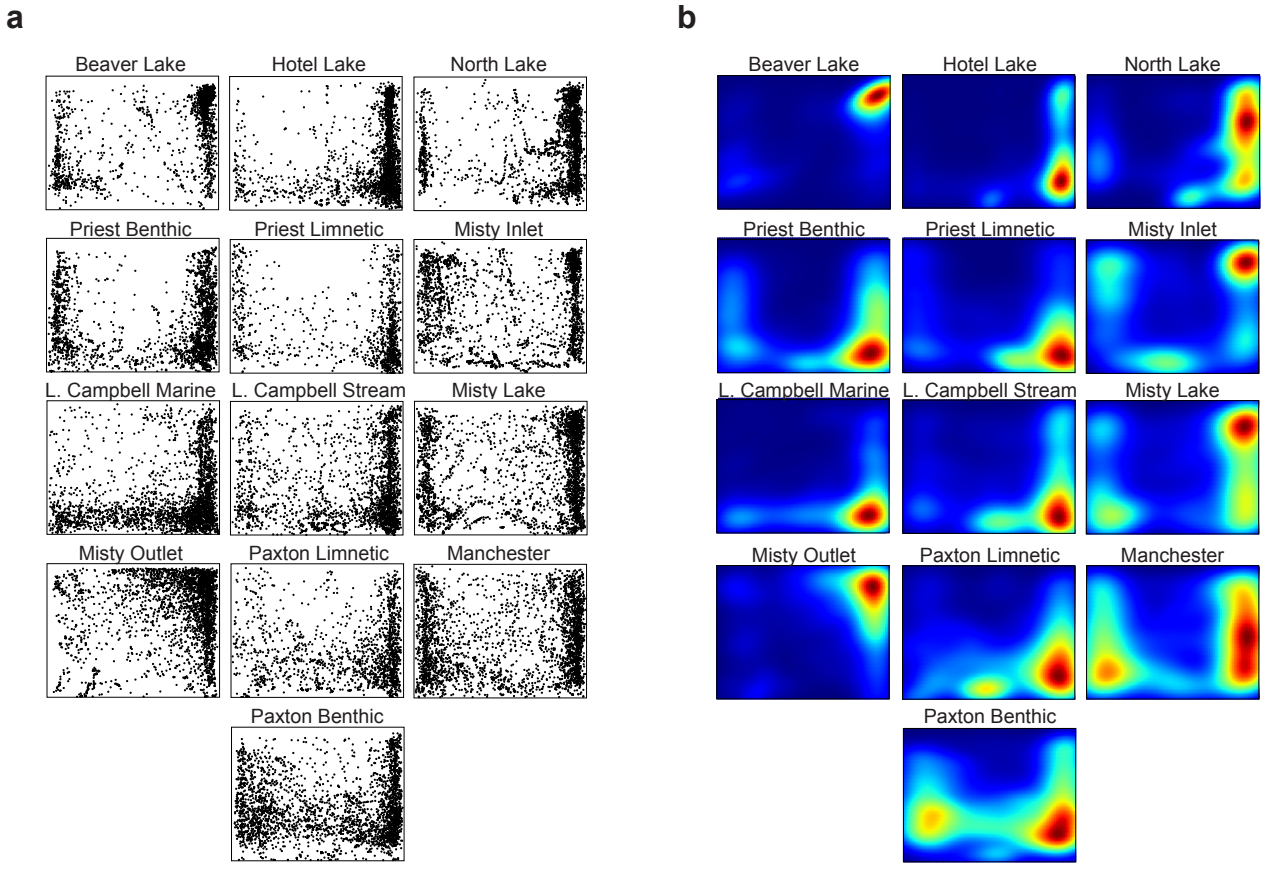
502

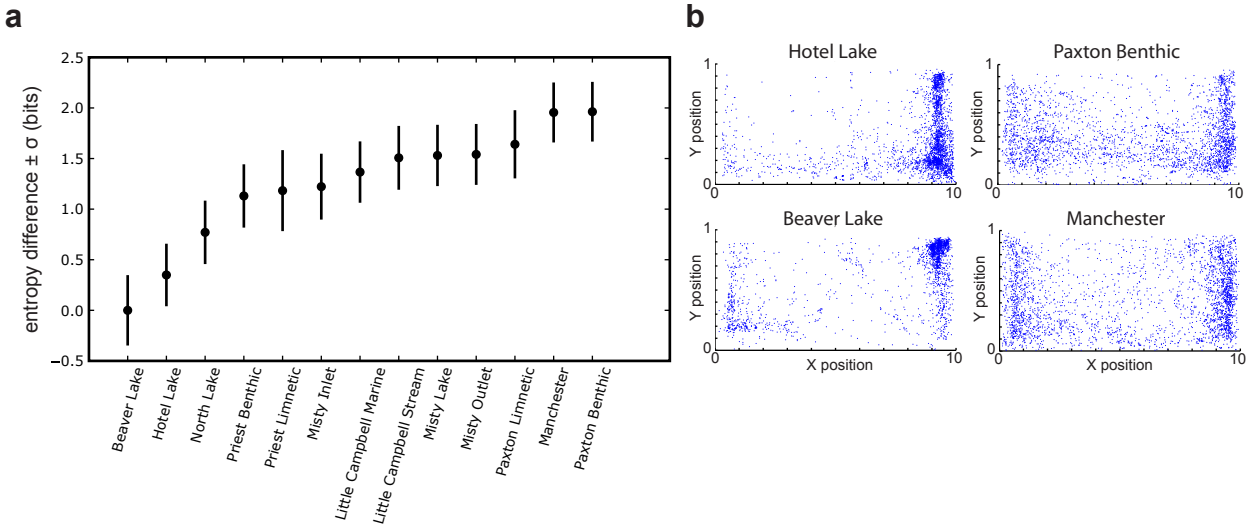
503 **Fig. 4** Population positional distributions are representative of individual behavioral patterns. **a** Tracks
504 of a single stickleback (hatched individual from (**b**)) superimposed on its population heat map. Red dots
505 indicate the starting position of the fish. **b** Log-likelihood analysis of the North Lake and Paxton Benthic

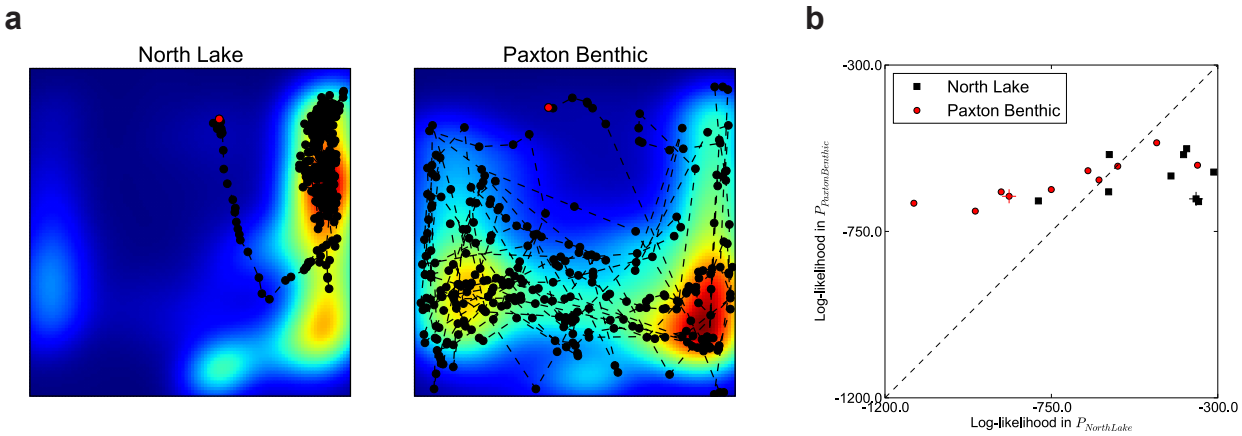
populations. Individual sticklebacks are represented by dots: North Lake = black squares; Paxton Benthic = red circles. Hatched individuals are featured in (**a**)

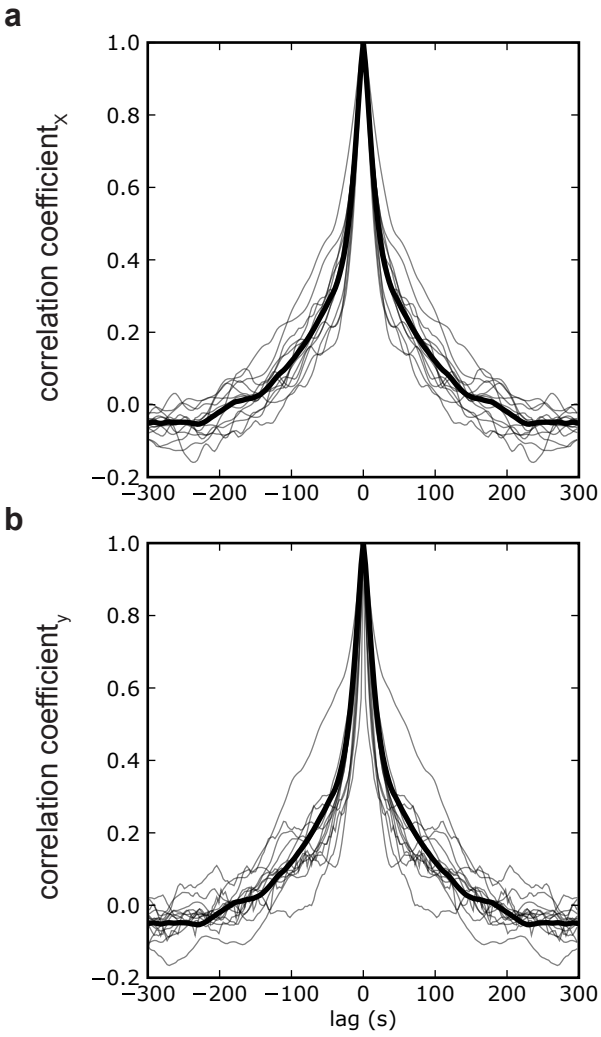
Fig. S1 Tracked fish positions are correlated in time in both horizontal and vertical dimensions for approximately 60 seconds, the width of the correlation peak at half-maximum. **a** Auto-correlation function shows correlation coefficient between each fish's horizontal position in the tank and that fish's horizontal position at a given time lag (thin grey lines). Average auto-correlation across all fish is shown in bold. **b** Auto-correlation function shows correlation coefficient between each fish's vertical position in the tank and that fish's vertical position at a given time lag (thin grey lines). Average auto-correlation across all 113 fish is shown in bold


light
22 cm
11 cm
30 cm
14 cm
32 cm
10 cm
54 cm
10 cm


a
Beaver Lake
Hotel Lake
North Lake
Priest Benthic
Priest Limnetic
Misty Inlet
L. Campbell Marine
L. Campbell Stream
Misty Lake
Misty Outlet
Paxton Limnetic
Manchester
Paxton Benthic
b
Beaver Lake
Hotel Lake
North Lake
Priest Benthic
Priest Limnetic
Misty Inlet
L. Campbell Marine
L. Campbell Stream
Misty Lake
Misty Outlet
Paxton Limnetic
Manchester
Paxton Benthic


a
b
entropy difference ± σ (bits)
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
Beaver Lake
Hotel Lake
North Lake
Priest Benthic
Priest Limnetic
Misty Inlet
Little Campbell Marine
Little Campbell Stream
Misty Lake
Misty Outlet
Paxton Limnetic
Manchester
Paxton Benthic
Hotel Lake
Paxton Benthic
Beaver Lake
Manchester
Y position
X position
0
1
10


a
North Lake
Paxton Benthic
b
-300.0
-750.0
-1200.0
North Lake
Paxton Benthic
Log-likelihood in $P_{PaxtonBenthic}$
Log-likelihood in $P_{NorthLake}$


a
b
correlation coefficient$_x$
correlation coefficient$_y$
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−300
−200
−100
0
100
200
300
lag (s)

		Horizontal (*x*) distribution										Vertical (*y*) distribution						Normality			
		Mean pos.		Med. pos.		Time		Ratio		sd		Mean pos.		Med. pos.		sd		*n*		$p_{(population)}$	
Population	*n*	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	*x*	*y*	*x*	*y*
BL	7	7.29	3.11	7.26	3.53	0.61	0.36	0.50	0.54	1.30	1.03	0.64[a]	0.20	0.65[a]	0.25	0.17[a]	0.06	1	0	*<1 e-10*	*<1 e-10*
HL	10	8.25	1.05	8.89	0.76	0.63	0.21	0.60	0.24	1.64	0.99	0.36[b]	0.15	0.32[b]	0.18	0.21	0.07	0	0	*<1 e-10*	*<1 e-10*
LM	10	7.06	1.77	7.39	2.01	0.30	0.20	0.25	0.25	2.00	1.10	0.27[b,e]	0.11	0.20[b,c]	0.10	0.20	0.08	0	0	*<1 e-10*	*<1 e-10*
LS	9	7.05	1.24	8.10	1.44	0.38	0.21	0.33	0.24	2.53	0.63	0.33[b]	0.15	0.26[b,h]	0.19	0.23	0.03	0	0	*<1 e-10*	*<1 e-10*
MC	10	5.71	2.70	5.56	3.35	0.33	0.28	0.15	0.43	2.39	1.13	0.44[c]	0.11	0.42[e]	0.15	0.23	0.05	3	0	*3 e-10*	*<1 e-10*
MI	9	5.63	3.01	6.07	3.54	0.36	0.33	0.23	0.46	1.83	0.80	0.50[f]	0.11	0.55[d,g]	0.23	0.28[b]	0.05	0	1	*0.00075*	*0.00120*
ML	10	6.16	2.67	6.23	3.69	0.40	0.22	0.22	0.42	2.44	0.94	0.46	0.15	0.49[d]	0.22	0.24	0.06	1	2	*<1 e-10*	0.56
MO	10	7.41	2.08	7.83	2.20	0.39	0.24	0.36	0.28	1.55	0.63	0.64[a,d]	0.14	0.70[a,f]	0.13	0.22	0.09	0	0	*<1 e-10*	*<1 e-10*
NL	9	7.53	2.53	8.02	2.82	0.50	0.25	0.40	0.50	1.42	0.82	0.47	0.19	0.45	0.22	0.18[a]	0.04	0	0	*<1 e-10*	*<1 e-10*
PB	10	5.76	2.33	5.98	2.90	0.29	0.27	0.19	0.33	2.33	1.04	0.39[b]	0.10	0.36[b]	0.10	0.20	0.04	1	0	*3.5 e-8*	*<1 e-10*
PL	6	6.62	1.90	7.08	2.37	0.35	0.31	0.29	0.35	2.27	0.96	0.31[b]	0.05	0.26[b]	0.04	0.23	0.03	1	0	*<1 e-10*	*<1 e-10*
RB	9	6.45	3.25	6.37	3.95	0.34	0.27	0.19	0.48	1.50	0.84	0.33[b]	0.11	0.27[b,h]	0.12	0.22	0.06	0	1	*<1 e-10*	*<1 e-10*
RL	4	6.88	1.40	7.98	1.41	0.45	0.22	0.31	0.32	3.06	0.73	0.30[b]	0.10	0.23[b]	0.11	0.23	0.05	1	0	*<1 e-10*	*<1 e-10*
Total	113	6.74	2.37	7.09	2.86	0.40	0.27	0.31	0.39	1.98	0.98	0.42	0.17	0.41	0.22	0.22	0.06	8	4	n/a	n/a

MANOVA								
A	Wilks Lambda < *1 e-10*							
$F_{(12,100)}$	1.081	1.215	1.556	1.088	2.396	6.851	7.674	2.109
p	0.384	0.284	0.117	0.378	*0.009*	*6.9 e-9*	*6.6 e-10*	*0.023*

Kruskal-Wallis								
$X^2_{(12)}$	14.728	11.733	16.595	16.154	23.693	48.610	51.747	23.137
p	0.257	0.467	0.165	0.184	*0.022*	*<1 e-10*	*<1 e-10*	*0.027*

Table 1 Summary data for thirteen populations tested in the shoaling assay. For horizontal distributions, mean position, median position, time shoaling, edge-corrected shoaling time ratio, and standard deviation are summarized for all thirteen populations. Similarly, for vertical

distributions, mean position, median position, and standard deviation are summarized. MANOVA and Kruskal-Wallis tests were used to compare populations in all shoaling measures. Test statistics and *p* values are displayed for each test, with significant *p* values in italic font. For MANOVA, Tukey's post hoc tests revealed pairwise differences between populations in vertical distribution measures. Superscript letters indicate significant pair-wise differences (a≠b, c≠d, e≠f, g≠h). Results of normality tests are presented at the far right. The number of individuals (*n*) whose distributions fail to deviate from normality are shown for each population in both horizontal (*x*) and vertical (*y*) dimensions. The significance scores for normality of whole population distributions ($p_{(population)}$) are also shown.