

# Bayesian analysis improves experimental studies about temporal patterning of aggression in fish



Eurico Mesquita Noleto-Filho<sup>a,b,\*</sup>, Ana Carolina dos Santos Gauy<sup>a,b</sup>, Maria Grazia Pennino<sup>c,d,e</sup>, Eliane Gonçalves-de-Freitas<sup>a,b</sup>

<sup>a</sup> Universidade Estadual Paulista Júlio Mesquita Filho (UNESP/IBILCE), Zoology and Botany Department, R. Cristóvão Colombo, 2265, CEP 15054-000, São José do Rio Preto, SP, Brazil

<sup>b</sup> Aquaculture Center of Sao Paulo State University (CAUNESP), Brazil

<sup>c</sup> Fishing Ecology Management and Economics (FEME), Universidade Federal do Rio Grande do Norte – UFRN. Depto. de Ecologia, Natal, RN, Brazil

<sup>d</sup> Statistical Modeling Ecology Group (SMEG), Departament d'Estadística i Investigació Operativa, Universitat de València, C/Dr. Moliner 50, Burjassot, 46100 Valencia, Spain

<sup>e</sup> Instituto Español de Oceanografía, Centro Oceanográfico de Murcia, C/Varadero 1. San Pedro del Pinatar, 30740, Murcia, Spain

## ARTICLE INFO

### Keywords:

Aggressive behavior  
Bayesian analysis  
Longitudinal design

## ABSTRACT

This study aims to describe a Bayesian Hierarchical Linear Model (HLM) approach for longitudinal designs in fish's experimental aggressive behavior studies as an alternative to classical methods. In particular, we discuss the advantages of Bayesian analysis in dealing with combined variables, non-statistically significant results and required sample size using an experiment of angelfish (*Pterophyllum scalare*) species as case study. Groups of 3 individuals were subjected to daily observations recorded for 10 min during 5 days. The frequencies of attacks, displays and the total attacks (attacks + displays) of each record were modeled using Monte Carlo Markov chains. In addition, a Bayesian HLM was performed for measuring the rate of increase/decrease of the aggressive behavior during the time and to assess the probability of difference among days. Results highlighted that using the combined variable of total attacks could lead to biased conclusions as displays and attacks showed an opposite pattern in the experiment. Moreover, depending of the study, this difference in pattern can happen more clearly or more subtly. Subtle changes cannot be detected when p-values are implemented. On the contrary, Bayesian methods provide a clear description of the changes even when patterns are subtle. Additionally, results showed that the number of replicates (15 or 11) invariant the study conclusions as well that using a small sample size could be more evident within the overlapping days, that includes the social rank stability. Therefore, Bayesian analysis seems to be a richer and an adequate statistical approach for fish's aggressive behavior longitudinal designs.

## 1. Introduction

In aquaculture, several species of fish are social and interact aggressively to set a dominance hierarchy and for territorial defense. Aggressive behavior can be showed in a very dynamic way, escalating in minutes, hours and days. These dynamic processes can cause injuries and, if continued for long time, lead to individual death, or to increased social stress that, in turn, cause negative effect on growth and general welfare (Damsgård and Huntingford, 2012). Hence, knowledge about the temporal dynamic of aggressive behavior can benefits fish's management, enhancing species growth and welfare (Dow et al., 1976; Gonçalves-de-Freitas et al., 2008; Damsgård and Huntingford, 2012;

Montero et al., 2009). Indeed, bibliography of studies intending to evaluate aggressive behavior in fish is increasingly growing (Gómez-Laplaza and Morgan, 1993; Ros et al., 2006; Almazán-Rueda et al., 2004; Castro and Caballero, 2004; Balzarini et al., 2014). In these articles, ethograms are usually used to understand how fish aggressively interact and to standardize variables patterns for quantifying the number of aggressive responses. Some typical examples of these variables are mouth attacks, lateral threats, and tail beating that can be used individually, partially combined based on its intensity (e.g. attacks and displays), or fully combined in a single one (e.g. total attacks) (Hallen and Wittenberger, 1987; Gomez-Laplaza and Morgan, 1993; Werneyer et al., 2002; Desjardins et al., 2012; Balzarini et al., 2014).

\* Corresponding author at: Universidade Estadual Paulista Júlio Mesquita Filho (UNESP/IBILCE), Zoology and Botany Department, R. Cristóvão Colombo, 2265, CEP 15054-000, São José do Rio Preto, SP, Brazil.

E-mail addresses: [euriconoleto@hotmail.com](mailto:euriconoleto@hotmail.com) (E.M. Noleto-Filho), [ana.gauy@gmail.com](mailto:ana.gauy@gmail.com) (A.C. dos Santos Gauy), [graziapennino@yahoo.it](mailto:graziapennino@yahoo.it) (M.G. Pennino), [elianegfreitas@gmail.com](mailto:elianegfreitas@gmail.com) (E. Gonçalves-de-Freitas).

<http://dx.doi.org/10.1016/j.beproc.2017.09.017>

Received 8 March 2017; Received in revised form 16 August 2017; Accepted 26 September 2017

Available online 29 September 2017

0376-6357/ © 2017 Elsevier B.V. All rights reserved.

Combining variables have become a common practice among fish aggressive behavior studies (Gonçalves-de-Freitas et al., 2008; Maan et al., 2001; Ros et al., 2006; Almazán-Rueda et al., 2004; Castro and Caballero, 2004), although there is still no clear the mathematical criteria about when and how do it.

However, several quantitative problems arise when aggressive behavior data are analyzed using both individual and combined variables, because the collection of this information usually presents three main issues: (i) replicates are hard to produce; (ii) the way which these behaviors occur is quite variable within individuals and; (iii) the experimental design is longitudinal (i.e. concerns the same individual over time). The use of classical approaches to account these issues could be problematic as it imposes difficulties in respecting classical parametrical assumptions as normality, homoskedasticity, and sphericity. Additionally, multiple comparisons are usually performed using post-hoc tests (e.g. Terleph, 2004; Carvalho et al., 2012). Nevertheless, when multiple comparisons are implemented, the probability that a researcher wrongly conclude that there is at least one statistically significant effect across a set of tests, increases additively with each test (Gelman et al., 2012). In addition, the evaluation of repeated measures when there are more than one treatment/independent variables and the data are non-parametrical is still a limitation in classical approaches.

The Repeated Measures Analysis of Variance (RM ANOVA) is often used in fish's aggressive behavior studies (e.g. Gómez-Laplaza and Morgan, 1993; Ros et al., 2006; Almazán-Rueda et al., 2004; Castro and Caballero, 2004). Alternatively, the use of Bayesian Hierarchical Linear Models (Bayesian HLM) is grown in popularity highlighting its usefulness in many biological fields including behavioral studies (Alterovitz et al., 2007; McCarthy, 2007; Kinas and Andrade, 2007; McNamara et al., 2006). Bayesian HLMs operate in a multilevel way making comparisons among groups and examining the posterior distributions from different perspectives or margins (Kruschke, 2015). Hence, they are more suitable for dealing with multiple comparison problems, because they shift point estimates and their corresponding intervals closer to each other instead of making confidence intervals wider as in classical estimates corrections (Gelman et al., 2012). Additionally, Bayesian HLMs provide a different view of the probability, defining it as a metric of uncertainty. On the contrary, the classical view of the probability is defined as the limit of the relative frequency of observed data in an experiment if the same procedure of data collection and analysis was implemented multiple times (McCarthy, 2007). For this reason, in a classical approach, we are not able to provide the exact probability of difference between two or more data groups as Bayesian methods do. Classical methods use p-values to access the probability of having an extreme result when the same experiment is repeated multiple times. Many researchers have pointed the limitations of using p-values in these cases as this approach ignores important non-significant results in the conclusions of the studies (Gelman, 2013; Ludwig, 2005; Pitak-Arnnop et al., 2010).

Bayesian analysis can work with resampling procedures, such as Monte Carlo Markov chains, which can numerically approximate Gaussian or different probability distributions (e.g. binomial, Poisson, etc) (Kruschke, 2012). Moreover, Bayesian methods can also implement heteroskedastic robust models that can overcome the assumption of homoskedasticity required in RM ANOVA (Kruschke, 2012), even with small samples sizes (Good and Zellner, 1975; Zellner, 1998). Indeed, as mentioned before, replicates in aggressive behavior studies can be difficult and expensive to be acquired. Consequently, the possibility to reduce sample size without changing the robustness of the conclusions could help scientists dealing with this typical issue.

Considering the importance of aggressive behavior studies for fish's welfare in aquaculture, knowledge on the longitudinal designs of the aggressive behavior is crucial for the proper understanding of aggression in fishes, and its consequences for fish social group as well. Within this context, for the first time, we discuss the advantages of Bayesian analysis in dealing with gathered data, non-statistically significant

results and required sample size. Therefore, this paper aims to test if a Bayesian HLM approach improves modeling longitudinal designs in aggressive behavior study in angelfishes, a social rank organized cichlid fish (e.g. Carvalho et al., 2012).

## 2. Materials and methods

### 2.1. Subjects and maintenance

The subject species was the angelfish, *Pterophyllum scalare* Schultze 1823. This fresh water cichlid presents a well known agonistic interaction and social rank organization (Gomez-Laplaza & Morgan 1993). The animals were provided by the Aquaculture Center (CAUNESP) of the São Paulo State University (UNESP) from Brazil. Individuals were kept for 15 days in water tanks of 500L polyethylene (ca. 1 fish/10L), and fed with a commercial ration developed for cichlids, twice a day, in the morning and evening (8:00 and 18:00), until apparent satiety. The water temperature was controlled to 28 °C (Pérez et al., 2003) and its quality was maintained through biological filters (filter 400L/h) and constant aeration. The photophase was kept on 12 h of light (07:00 h to 19:00 h). The tanks were siphoned weekly to remove the remnants of food and feces.

### 2.2. Experimental design

Fifteen groups consisting of three animals with size between 30 and 50 mm (standard length) were kept for 5 days to behavioral observation. In order to confirm consistency, this experiment was repeated twice. Aggressive behavior was daily recorded during 10 min. The records were run always between 14:00 h and 18:00 h to avoid influence of the circadian rhythm. This interval of time was chosen to avoid the influence of competition for food (Gómez-Laplaza and Morgan, 2003; Grobler and Wood, 2013). The aggressive interactions were quantified by the frequency of aggressive behavior, based on *P. scalare* ethogram (Carvalho et al., 2012). Data was labeled as attacks (more intensive, with contact), displays (less intensive, no contact), and total attacks (attacks + displays) for the analysis (Haller and Wittemberger, 1988; Alvarenga and Volpato, 1995). Body bites, mouth fights and chasing were considered as attacks, while lateral threat, frontal displays and ripples as displays.

### 2.3. Statistical modeling

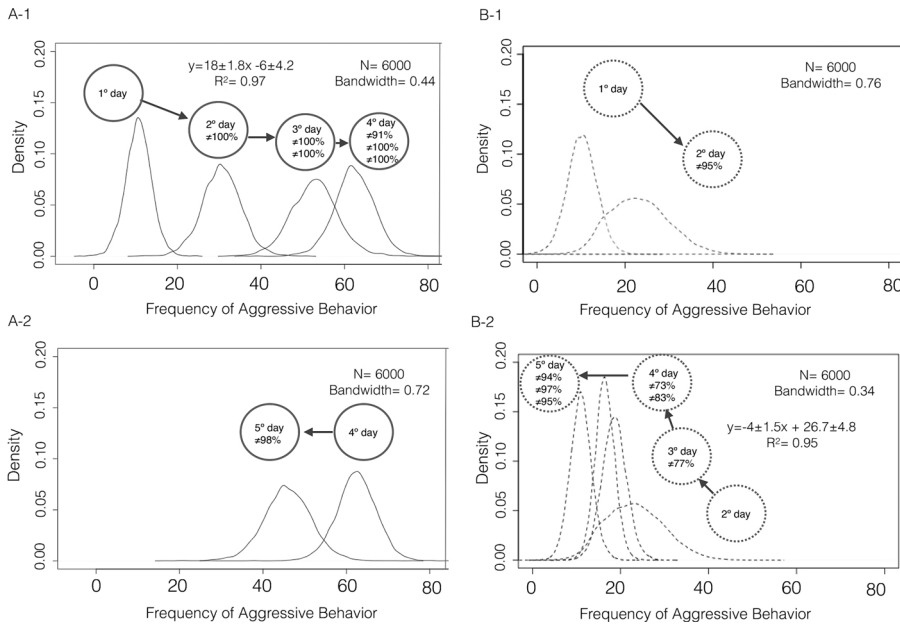
Modeling was done with data of attacks, displays and total attacks of each trial that were adjusted with Monte Carlo Markov chains (Metropolis-Hasting algorithm) using non-informative priors (Jeffreys, 1961).

#### 2.3.1. Modeling the rate of increase/decrease

Bayesian regression analysis was performed for measuring the rate of increase/decrease of the aggressive behavior per day (agB/day). The model was specified as:

$$\begin{aligned} W_i &\sim N(\mu_i, \tau_i) \\ \mu_i &= \alpha + bY_i \\ \alpha &\sim N(\sigma, t) \\ b &\sim N(0, 0.10E - 6) \\ \sigma &\sim N(0, 0.10E - 6) \\ t &\sim \text{Gamma}(0.001, 0.001) \end{aligned} \quad (2.1)$$

where  $W_i$  indicates the error term for  $i$  observations with a normal distribution,  $\sigma$  and  $t$  are hyperparameters that specify random variances for the intercept  $\alpha$  and slope  $b$  of the linear model. Uninformative priors were assigned for all the parameters.



**Fig. 1.** The daily agB density distributions of the displays (A-1 and 2) and attacks (B-1 and 2) of the first trial. It was observed in 15 min at each day. Each circle contains the probability of difference from the days before. The first probability is from the first day before, the second is from second day before and so on.

### 2.3.2. Modeling the probability of difference

A Bayesian HLM was performed to access the probability of difference among days using the last day as the reference class (Wetzels et al., 2012). Subsequently, it was calculated the probability that treatments are equal, i.e. their difference is equal to 0. For this purpose, it was used a boolean variable that counts the number of simulations in which the difference among days are  $\geq 0$  or  $< 0$ , representing a Bayesian analogue to the classical approach p-value (Gelman et al., 1996; Kéry and Royle, 2016). In this case, the model was specified as:

$$\begin{aligned}
 W_i &\sim N(\mu_i, \tau[Y_i]) \\
 \mu_i &= \alpha + b[Y_i] \\
 b &\sim N(\sigma, t) \\
 \alpha &\sim N(0, 0.10E - 6) \\
 \sigma &\sim N(0, 0.10E - 6) \\
 t &\sim \text{Gamma}(0.001, 0.001) \\
 \tau &\sim \text{Gamma}(0.001, 0.001)
 \end{aligned} \quad (2.2)$$

where  $W_i$  indicates the error term for  $i$  observations with a normal distribution and  $b$  specifies the average for each category  $Y$  for the parameter  $\mu$ . The hyperparameters  $\sigma$  and  $t$  indicate random variances for the parameter  $b$  and  $\alpha$  parameter refers to the intercept (reference class) and is specified by an uninformative prior distribution. As in the precedent model, uninformative priors were assigned for all the parameters.

### 2.3.3. Reducing sample size

For measuring the feasibility of reducing the sample size, we performed a simulation study, using our data. We used the same model of the probability differences described above. However, we performed it multiple times, randomly taking one replicate away from the data each time. The Cohen d effect size (1988) was calculated to measure the impact of reducing the sample size in data. It was considered a value of 0.2, 0.5 and 0.8 for the small, medium and large effects, respectively (Cohen, 1988).

### 2.3.4. Plotting the results

A Bayesian model without any reference class parameters was adopted for plotting the results graphically. All the posterior distributions were plotted in parts for didactic purposes.

Here, the model was specified as:

$$\begin{aligned}
 W_i &\sim N(\mu_i, \tau[Y_i]) \\
 \mu_i &= b[Y_i] \\
 b &\sim N(\sigma, t) \\
 \sigma &\sim N(0, 0.10E - 6) \\
 t &\sim \text{Gamma}(0.001, 0.001)
 \end{aligned} \quad (2.3)$$

where  $W_i$  indicates the error term for  $i$  observations with a normal distribution and  $b$  indicates the average for each category  $Y$  for the parameter  $\mu$ . The hyperparameters  $\sigma$  and  $t$  specify random variances for the parameter  $b$ . Uninformative priors were assigned for all the parameters.

### 2.4. Software and simulations

Bayesian models were computed using the JAGS software (Just Another Gibbs Sampler) and statistical package “rjags” (Plummer et al., 2006), specifically developed for implement Bayesian methods using the Monte Carlo Markov chains through the R software (R Development Core Team, 2016). The treatments parameters were estimated through simulations. The number of simulations was established by trial and error in the model convergence process. The success or failure in convergence was checked using the criteria of Geweke (Geweke, 1992) and Gelman-Rubin (Gelman and Rubin, 1992). The latter was also used for establishing a period of “burn-in” constituted by the disposal of some chains to help the model convergence.

## 3. Results

### 3.1. Rate of increase/decrease and difference probability among days in fish's attacks and displays

For the first trial, the displays increased till the fourth day at a rate of  $21 \pm 3$  agB/day (Fig. 1A-1). This increase was very consistent since the posterior distributions from the first to the third day have not overlapped, being a 100% different from each other. At the fourth day, displays still increased, although this rate dropped to  $17.9 \pm 1.8$  agB/day. This drop caused an overlap between the distribution of the fourth and third day, differing in 91% from each other. At the fifth day, the agB remarkably decreased with a 98% difference from the fourth day (Fig. 1A-2). The attacks increased until the second day (not enough data to calculate a rate) (Fig. 1B-1) and retreated from the second to the fifth day at a rate of  $-4.1 \pm 1.5$  agB/day (Fig. 1B-2). Different from

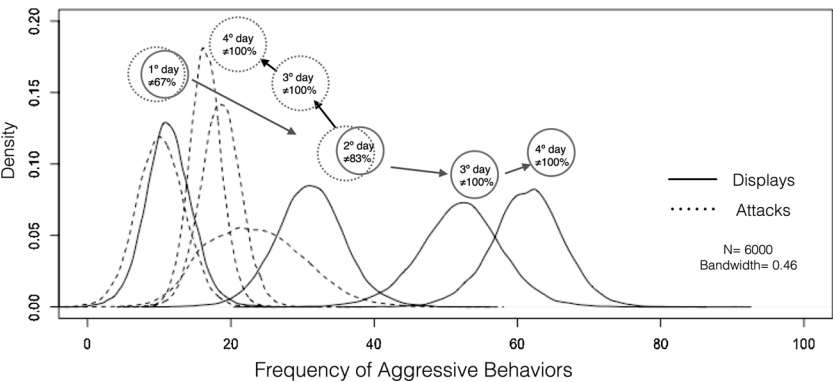


Fig. 2. The daily agB density distributions of the attacks and the displays of the first trial. It was observed in 15 min at each day. They are represented by solid and dashed lines, respectively. Each circle contains the probability of difference between attacks and displays at each day.

**Table 1**  
The mean and 95% credible interval for the standard deviation among and within days of the attacks and displays variables in two trials.

	Attacks		Displays	
	Mean	95% CI	Mean	95% CI
First Trial				
Among all days	13.95	5.27–36.95	30.51	13.02–71.81
Within 1° day	12	8.15–18.34	7.48	5.05–11.29
Within 2° day	21.26	14.54–32.12	24.2	16.65–36.68
Within 3° day	12.08	8.23–18.53	32.12	21.85–48.67
Within 4° day	9.86	6.76–14.67	35.79	24.37–54.37
Within 5° day	3.07	2.08–4.66	24.47	16.69–37.22
Second Trial				
Among all days	9.49	1.7–28.0	29.85	12.56–72.42
Within 1° day	13.31	9.0–20.2	12.53	8.58–18.92
Within 2° day	27.66	18.6–42.2	18.6	12.74–28.36
Within 3° day	11.2	7.6–17.1	22.34	15.18–34.21
Within 4° day	9.12	6.2–13.8	19.27	13.15–29.24
Within 5° day	9.63	6.5–14.6	22.91	15.56–34.66

displays, there was a high overlapping of the distributions from the second to the fourth day and an increase in difference at the fifth day. It is noteworthy that the attacks and displays were almost the same till the second day, differing in 83% (Fig. 2). However, at the third day, both behaviors became completely non overlapped till the rest of the study period. In addition, the credible interval for the standard deviation among days more than encompassed the standard deviations within each day (Table 1, Fig. 3A-1 and A-2), meaning that there is more uncertainty in differences among days than within each day for this trial.

For the second trial, the displays increased till the third day at a rate

of  $27.9 \pm 3.72$  agB/day (Fig. 4A-1). The second and third day overlapped in 8%. The displays decreased till the fifth day at a rate of  $-2.5 \pm 5$  agB/day (Fig. 4A-2). Although displays decreased at the fourth day, this day was very similar to the third and the fifth ones, differing only in 71% and 54%, respectively. The attacks increased from the first to the second day (not enough data to calculate a rate), decreasing till the fifth day at a rate of  $-8.6 \pm 1.7$  agB/day (Fig. 4B-1 and B-2). Different from the first trial, there was a high overlapping only from the second to the third day. At the fourth and fifth day, the decrease was clearer with a slightly overlapping of the distributions on the fourth day and a not overlapping at the fifth day. As in the first trial, the attacks and displays were the same at the first day ( $\neq 50\%$ ) (Fig. 5). Although, at the second trial displays and attacks were more likely to differ on the second day ( $\neq 94\%$ , compared to 83% in the first trial). Additionally, the rate of increase of displays (till the third day) and a decrease of attacks differed in 92% and 96% between trials, respectively. In addition, the credible interval for the standard deviation among days more than encompassed the standard deviations within each day, in most cases (Table 1, Fig. 3B-1 and B-2). The only exception was the second day of the variable attacks. Therefore, in most cases, there was more uncertainty in differences among days than within each day for this trial.

3.2. Rate of increase/decrease and difference probability among days in fish's total attacks

Overall, the displays were much higher than the attacks. Therefore, they prevailed on the total attacks. Since, attacks retreated and displays increased after the second day, gathering these variables, negatively impacted the rate of increase reducing it from  $18 \pm 1.8$  agB/day to  $17.3 \pm 3$  agB/day at the first trial (Figs. 6A-1 and Fig. 1) and from

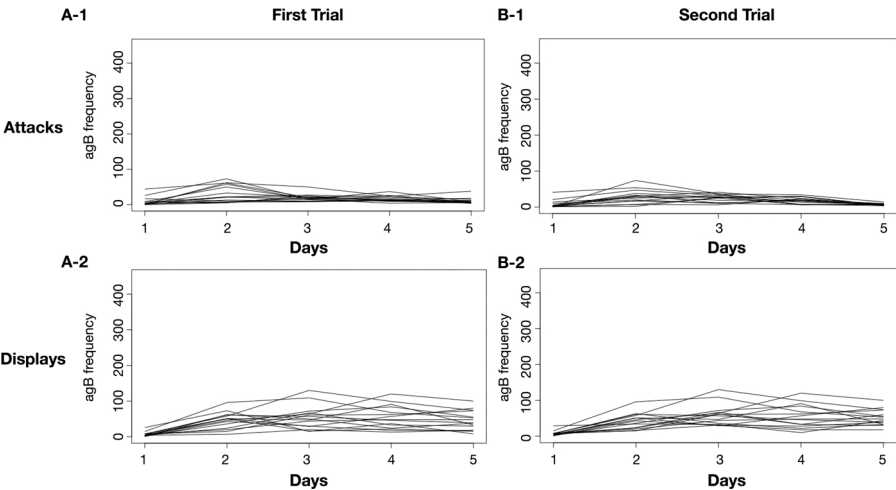
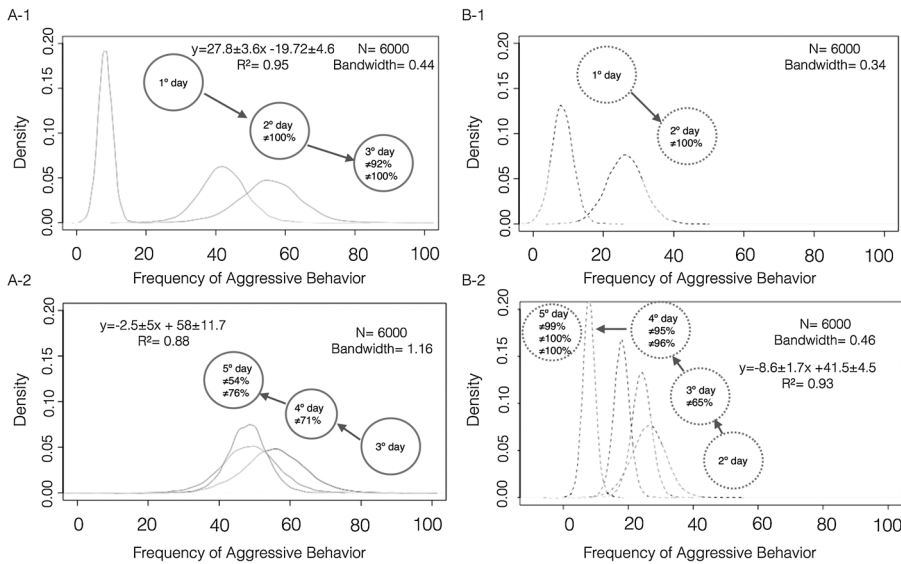


Fig. 3. The daily agB frequency of each of the fifteen groups in five days os observation. It was observed in 15 min at each day.





**Fig. 4.** The daily agB density distributions of the displays (A-1 and 2) and the attacks (B-1 and 2) of the second trial. It was observed in 15 min at each day. Each circle contains the probability of difference from the days before. The first probability is from the first day before, the second is from second day before and so on.

$27.8 \pm 3.6$  agB/day to  $24 \pm 8$  agB/day at the second (Fig. 6B-1 and Fig. 4). This rate reduction is due to the overlapping of the distributions from the days after the second one, that is the period in which attacks and displays followed different paths (Figs. 2 and 5). Hence, the difference between the second and third day, which is a 100% of displays, dropped to an 84% of total attacks in the first trial. At the second trial the difference of 92% at displays dropped to 79% at total attacks. The rate of increase between total attacks trials differed in 84%.

### 3.3. Reducing sample size

For displays of the first trials, the reduction of the data sample size tended to decrease the posterior probability of difference among all the days. The unique exception was found between the third and fourth day, in which the posterior probability did not present a clear pattern, also showing a small effect size (Table 2). The difference among the first three days remained the same till the tenth replicate ( $\neq 100\%$ ), and decreasing till the rest of the sample reduction (Fig. 7A-1). Additionally the effect size between these days was considered large (Table 2). For the attacks, the reduction of the data sample size tended to decrease the posterior probability of difference among the first and second day, and the fourth and fifth day. Between these days there was as medium effect size. From the second till the fourth day there was no clear pattern, with a small effect size (Fig. 7A-2 and Table 2).

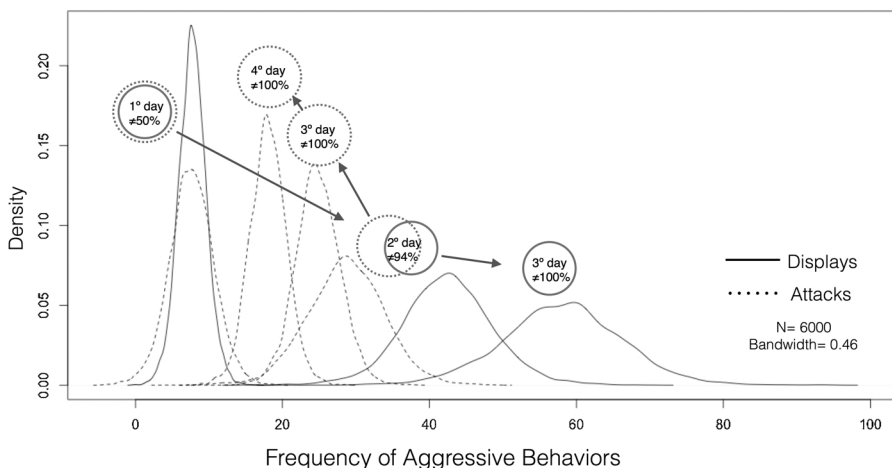
At the second trial, for displays, the sample size data reduction

tended to decrease the posterior probability of difference among the first three days (Fig. 7B-1). Although, different from the first trial, only the difference among the first and second day remained the same till the seventh replicate ( $\neq 100\%$ ). Between the first and second day there was a very large effect size, while among the second and third days there was a medium effect. For attacks reducing sample size provided similar patterns to the first trial (Fig. 7B-2), although, the first and last two days presented a large effect size instead of a medium one.

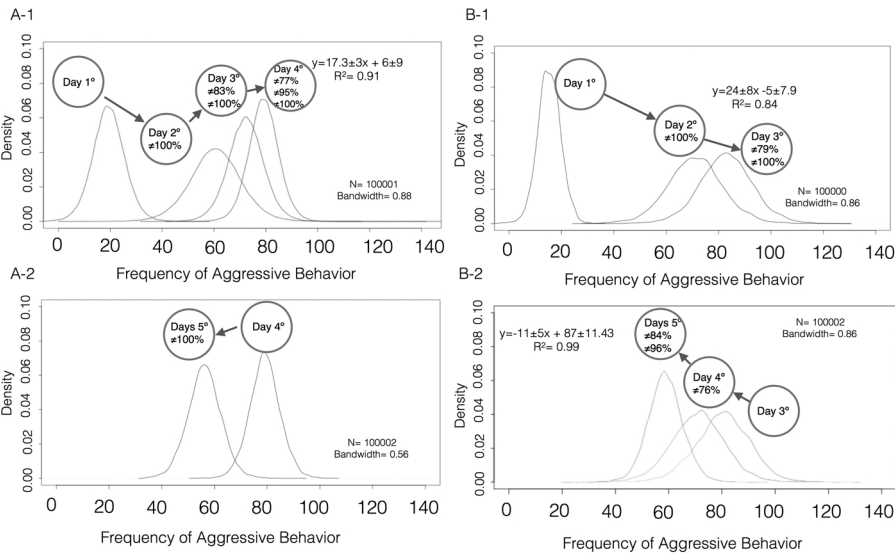
## 4. Discussion

In this study we described how Bayesian analysis could be a reliable alternative to classical methods for longitudinal designs in fish's experimental aggressive behavior studies. Particularly, we presented the advantages of Bayesian analysis in dealing with combined variables, non-statistically significant results and required sample size using as case study an experiment of angelfish (*Pterophyllum scalare*) species.

Our findings revealed how probability degrees, in accordance with the density distributions, provided a clear description of the changes even when patterns are subtle, which cannot be detected using p-values. In addition, Bayesian analysis is enabled to identify whether it is feasible or not to combine variables and reduce sample size. All these subjects are discussed in the topics below.



**Fig. 5.** The daily agB density distributions of the attacks and displays of the second trial. It was observed in 15 min at each day. They are represented by solid and dashed lines, respectively. Each circle contains the probability of difference between attacks and displays at each day.



**Fig. 6.** The daily agB density distributions of the total attacks (TA). A-1 and A-2 represents the first trial. B-1 and B-2 represents the second trial. It was observed in 15 min at each day. Each circle contains the probability of difference from the days before. The first probability is from the first day before, the second is from second day before and so on.

**Table 2**  
Effect sizes between days.

Days	Displays First Trial	Attacks First Trial	Displays Second Trial	Attacks Second Trial
1–2	1.27	0.64	2.00	1.00
2–3	1.00	0.23	0.50	0.12
3–4	0.45	0.20	0.10	0.56
4–5	0.77	0.60	0.60	0.90

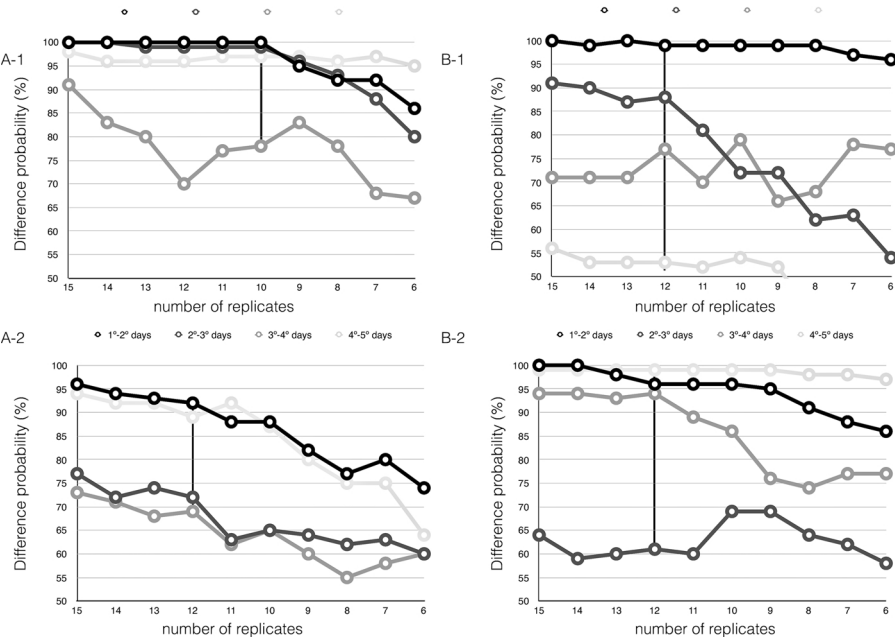
4.1. Behavioral consistency

A consistent pattern shows that a specific behavior does not occur by chance. The same experiment, repeated twice, shows a similar pattern for attacks, displays, and total attacks. Gómez-Laplaza and Morgan (1993) have also found similar patterns for *P. scalare*, which reinforce our results. The increase rate of total attacks, attacks, and displays, between the trials, was also similar (differed in 84%, 94%, and 92%, respectively) with the mentioned study. The uncertainty of 8 and 3 agB/day for total attacks, 3 and 3.72 agB/day for displays, and 1.5

and 1.8 agB/day for attacks was small. We can conclude that the rate of increase is almost similar within the replicates of each trial. In addition, the uncertainty of the standard deviation among days was bigger than within days, meaning that most differences are due to time rather than the unknown factors. It is noteworthy, that the second day presented the larger variation within days, for the variable attacks. For the second trial, the variation within this day was even higher than between days. The second day is related to the establishment of the dominance and subordinate relationships within groups, which is related to the increase in high-intensity agonistic behaviors (Ros et al., 2006). There are many known factors which can increase the variability of this social hierarchy, such as size, experience and physiological status (Damsgård and Huntingford, 2012). Hence, this high variation within the second day is probably related to the social rank stability.

4.2. Statistical significance

Aggressive behaviors are subject to changes during the time (Dow et al., 1976). However, these changes can happen so subtly that may not be statistically noticed in a short period of time. For example, in



**Fig. 7.** The distribution of difference between days with different sample sizes. A-1 and 2 represents the displays and attacks of the first trial. B-1 and B-2 represents the displays and attacks of the second trial. The black line is between the first and second day, the dark gray line is between the second and third days, the gray is between the third and fourth days and the light is between the fourth and fifth days. Black line is the cut point.

total attacks the second day is 83% different from third and 95% different from day fourth, but the third day is 77% different from the fourth day (Fig. 6A-1). Using uninformative priors, Bayesian analysis may provide results similar to classical approaches (Jaynes, 1976). When we find 95% of difference, probably we also would find significance with the classical approach ( $p < 0.05$ ). In this case, between days 2 and 3 and 3 and 4 there was no statistical significance considering a 95% credible interval. On the contrary, a statistical difference was found between days 2 and 4. Usually, when this happens, researchers state that the aggressive behavior is the same between the non-statistically significant days even when there is an increase/decrease pattern (e.g. Gómez-Laplaza and Morgan, 1993; Adams et al., 2000; Terleph, 2004; Gonçalves de Freitas and Mariguela, 2006). Indeed when considering p-values we might misinterpret non-significant results (Gelman, 2013; Ludwig, 2005; Pitak-Arnop et al., 2010). Bayesian analysis clarifies the longitudinal changes by building the agB distribution in each day and calculating the probability of difference between them. Therefore, we can measure the strength of the evidence against the null hypothesis (days are equal) using probability degrees (Wagenmakers, 2007). A very intuitive classification was proposed by Raftery (1995) where the categories 50% to 75%, 75% to 95%, 95% to 99% and  $> 99\%$  indicates weak, positive, strong and very strong evidence against the null hypothesis, respectively. At the total attacks example, the probability that the third day differs from the second is 83%, and it increases to 95% on the fourth day. Hence, the third day is a transition from second day to fourth day, going from a positive to strong evidence against the null hypothesis. Therefore, the Bayesian approach provides an idea of changes and continuity among days even with the lack of statistical significance, which is more intuitive for a longitudinal design than the equality idea provided by classical methods.

#### 4.3. Combining variables

As the numbers of less intensive aggressive behaviors (displays) were much higher than the numbers of the more intensive ones (attacks), they prevailed on the total attacks, shaping its distribution through time. In the literature, there are many articles using total attacks (e.g. Almazán-Rueda et al., 2004; Wessel et al., 2006; Castro and Caballero, 2004; Earley et al., 2006; Miyai et al., 2011; Carvalho et al., 2012; Pinho-Neto et al., 2014). However, our results highlight that adding attacks to displays diminishes the posterior probability between the days compared to using displays alone, losing their statistical significance considering a 95% credible interval (Figs. 1 A-2 and 5 A-1). This is because attacks and displays are different things in fish's social environment. Attacks (which require more energy) retreat in time when dominance hierarchy is established and displays (requires less energetic cost) increases (Haller and Wittenberger, 1988; Maan et al., 2001; Ros et al., 2006).

This pattern is also described for other species such as juvenile *Oreochromis niloticus* (Alvarenga and Volpato, 1995), *Tilapia zillii* (Neat et al., 1998) and *Oreochromis massambicus* (Ros et al., 2006). Although, this pattern is not clear for all species, as in hatchery-reared steelhead (*Oncorhynchus mykiss*) (Berejikian et al., 2001) and oscar (*Astronotus ocellatus*) (Gonçalves-de-Freitas and Mariguela, 2006). In this sense, whether or not using total attacks may be species-specific and should be previously investigated. In this context, Bayesian analysis can assess the impacts of combining data as attacks and displays by exhibiting the changes in probabilities that happens when using a variable alone and combined. Researchers could use the Bayesian tool to analyze the impact of combining the same variables for other species or other aggressive variables such as bites, chasing or tail beats. Considering the Raftery (1995) rule of thumb any 5% of difference at the posterior probability against the null hypothesis is enough to change from strong (95% to 99%) to positive evidence (75% to 95%). Therefore, we suggest a 5% change in the posterior distribution as a criterion to not

combining data. It is noteworthy that the best procedure will be always to compare the combined variable with the more quantitative single variable (in our example displays). Consequently, the expected effect is always negative or nonexistent since the predominant variable is the main one at the posterior distribution of the gathered variable. The same issue could be differently investigated with classical approaches checking if p-values are  $> 0.05$  or  $< 0.05$  when using the data gathered or not. However, even using this approach the amount of probability change would not be measured. Indeed, if for example, with gathered data the p-value will be 0.01 and with not gathered data of 0.06, this does not mean that using gathered data decreases in 5% the data difference (that is what Bayesian analysis can conclude). On the contrary, it decreases by 5% the probability of having an extreme result if the same experiment is repeated multiple times. Therefore, classical approaches did not measure the impact of combining in our experiment, but the effect of repeating it. Additionally, since classical approaches are mainly focused on p-values and statistical significance, it would not be applied to non-statistically significant results and these can be important in aggressive behavior longitudinal studies as explained in the topic above.

#### 4.4. Reducing sample numbers

The impact of using a small sample size was more evident in days that overlapped more. The inference among these days presented smaller effect sizes. When the effect size of the intervention is large, it is possible to detect such effect in smaller sample numbers, whereas a smaller effect size would require larger sample size (Sullivan and Feinn, 2012). Therefore, attacks were more sensitive to sample numbers from the second till the fourth days, given the smaller effect size and high overlapping among agB distribution. The displays were more sensitive among the third and the fourth day, which are the climax of increase for the first and second trials, respectively. Therefore this sensitivity is due to the uncertainty of the display's increase climax, which can happen at the third or fourth day depending on the individual. Besides the increase in variability, the expected pattern of increase of the agB for displays and decrease for attacks remained robust till the eleventh and tenth replicates, at the first and second trials, respectively. This pattern is related to the social rank stability (Alvarenga and Volpato, 1995; Neat et al., 1998; Ros et al., 2006). After the tenth replicate the variability reaches a level of increase that causes the misleading idea that the agB distribution of some days are equal. Hence, a small sample size could provide the false idea that the agB does not change after dominance hierarchy being established. As previously mentioned, when we rely on p-values with 95% intervals we ignore important no significant information. Thus, the sample size is important to acknowledge the transitional pattern of the agB among days. Therefore a small sample number ( $\leq 10$ ) can ignore these subtle changes, wrongly transmitting the idea that aggressive behaviors are the same at some days, and even Bayesian estimation can be biased.

The social rank stability is related to this pattern of decrease of the attacks and increase of the displays (Alvarenga and Volpato, 1995; Neat et al., 1998; Ros et al., 2006). Hence a small sample size could provide the false idea that the agB did not change after dominance hierarchy being established.

Usually articles concerning sample size uses power rather than effect size. Effect size and power are related. However, power considers a 95% interval and we are interested at differences that are below this interval. Therefore, we have chosen to work only with the effect size, since it fits better to our proposal.

## 5. Conclusions

For *P. scalare* the use of total attacks could lead to biased conclusions as displays and attacks showed an opposite pattern in the experiment. The daily changes in aggressiveness happens subtly and

cannot be detected when p-values are implemented. On the contrary, Bayesian methods provide a clear description of the changes even when patterns are subtle. Additionally, results showed that the number of replicates (15 or 11) invariantly the study conclusions as well that using a small sample size could be more evident within the overlapping days, that includes the social rank stability. Therefore, we can conclude that the Bayesian approach could be a richer and an adequate statistical framework for fish's aggressive behavior longitudinal designs.

### Funding

Eurico M. Noleto Filho – fellowship from CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*). Ana C.G. Santos – fellowship from FAPESP (Fundação e Amparo à Pesquisa do Estado de São Paulo – #2013/09021-0). Eliane Gonçalves-de-Freitas – grants from CNPq – #2016-26160-2

### Ethical Note

This study was carried out in accordance with the ethical principles of the National Council for the Control of Animal Experimentation (CONCEA, Brazil) and was approved by the Ethics Committee for Animal Experimentation of the São José do Rio Preto campus of São Paulo State University, known locally as the Universidade Estadual Paulista (UNESP), in Brazil (permit number 077/2013). All fish handling was preceded by anesthesia, and substantial efforts were made to minimize animal suffering, as recommended by the Animal Behavior Society's Guidelines.

### Acknowledgments

We acknowledge the entire team of the laboratory of Animal Behavior from the State University of São Paulo, Brazil, for helping with the samplings.

### References

- Adams, C., Huntingford, F., Turnbull, J., Arnott, S., Bell, A., 2000. Size heterogeneity can reduce aggression and promote growth in Atlantic salmon parr. *Aquac. Int.* 8, 543–549. <http://dx.doi.org/10.1023/A:1009255612529>.
- Almazán-Rueda, P., Schrama, J.W., Verreth, J.A.J., 2004. Behavioural responses under different feeding methods and light regimes of the African catfish (*Clarias gariepinus*) juveniles. *Aquaculture* 231, 347–359. <http://dx.doi.org/10.1016/j.aquaculture.2003.11.016>.
- Alterovitz, G., Liu, J., Afkhami, E., Ramoni, M.F., 2007. Bayesian methods for proteomics. *Proteomics* 7, 2843–2855. <http://dx.doi.org/10.1002/prot.200700422>.
- Alvares, C.M.D., Volpato, G.L., 1995. Agonistic profile and metabolism in alevis of the Nile tilapia. *Physiol. Behav.* 57, 75–80. [http://dx.doi.org/10.1016/0031-9384\(94\)00206-K](http://dx.doi.org/10.1016/0031-9384(94)00206-K).
- Balzarini, V., Taborsky, M., Wanner, S., Koch, F., Frommen, J.G., 2014. Mirror, mirror on the wall: The predictive value of mirror tests for measuring aggression in fish. *Behav. Ecol. Sociobiol.* 68, 871–878. <http://dx.doi.org/10.1007/s00265-014-1698-7>.
- Berejikian, B.A., Tezak, E.P., Riley, S.C., LaRae, A.L., 2001. Competitive ability and social behaviour of juvenile steelhead reared in enriched and conventional hatchery tanks and a stream environment. *J. Fish Biol.* 59, 1600–1613. <http://dx.doi.org/10.1006/jfbi.2001.1789>.
- Carvalho, T.B., Ha, J.C., Gonçalves-de-Freitas, E., 2012. Light intensity can trigger different agonistic responses in juveniles of three cichlid species. *Mar. Freshw. Behav. Physiol.* 45, 91–100. <http://dx.doi.org/10.1080/10236244.2012.690564>.
- Castro, J.J., Caballero, C., 2004. Effect of the light intensity upon the agonistic behaviour of Juvenile of white-seabream (*Diplodus sargus cadenati* de la Paz, Bauchot and Daget, 1974). *Aggress. Behav.* 30, 313–318. <http://dx.doi.org/10.1002/ab.20023>.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. Stat. Power Anal. Behav. Sci. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Desjardins, J.K., Hofmann, H., Fernald, R.D., 2012. Social context influences aggressive and courtship behavior in a cichlid fish. *PLoS One* 7. <http://dx.doi.org/10.1371/journal.pone.0032781>.
- Damsgård, B., Huntingford, F., 2012. Fighting and aggression. *Aquaculture and Behavior*. pp. 248–285. <http://dx.doi.org/10.1002/9781444354614.ch9>.
- Dow, M., Ewing, A.W., Sutherland, I., 1976. Studies on behavior of cyprinodont fish 3. Temporal patterning of aggression in apyosemion-striatum (Boulenger). *Behaviour* 59, 252–268. <http://dx.doi.org/10.1163/156853976X00398>.
- Earley, R.L., Edwards, J.T., Aseem, O., Felton, K., Blumer, L.S., Karom, M., Grober, M.S., 2006. Social interactions tune aggression and stress responsiveness in a territorial cichlid fish (*Archocentrus nigrofasciatus*). *Physiol. Behav.* 88, 353–363. <http://dx.doi.org/10.1016/j.physbeh.2006.04.002>.
- Gómez-Laplaza, L.M., Morgan, E., 1993. Social isolation, aggression, and dominance in attacks in juvenile angelfish, *Pterophyllum scalare*. *Aggress. Behav.* 19, 213–222. [http://dx.doi.org/10.1002/1098-2337\(1993\)19:3<213::AID-AB2480190306>3.0.CO;2-X](http://dx.doi.org/10.1002/1098-2337(1993)19:3<213::AID-AB2480190306>3.0.CO;2-X).
- Gómez-Laplaza, L.M., Morgan, E., 2003. The influence of social rank in the angelfish, *Pterophyllum scalare*, on locomotor and feeding activities in a novel environment. *Lab. Anim.* 37, 108–120. <http://dx.doi.org/10.1258/00236770360563741>.
- Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472.
- Gelman, A., Meng, X.-L., Stern, H.S., 1996. Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Stat. Sin.* 6, 733e807.
- Gelman, A., Hill, J., Yajima, M., 2012. Why we (usually) don't have to worry about multiple comparisons. *J. Res. Educ. Eff.* 5, 189–211. <http://dx.doi.org/10.1080/19345747.2011.618213>.
- Gelman, A., 2013. P values and statistical practice. *Epidemiology* 24, 69–72. <http://dx.doi.org/10.1097/EDE.0b013e31827886f7>.
- Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Stat.* 4, 169–193.
- Gonçalves-de-Freitas, E., Mariguela, T.C., 2006. Social isolation and aggressiveness in the Amazonian juvenile fish *Astronotus ocellatus*. *Braz. J. Biol.* 66, 233–238. <http://dx.doi.org/10.1590/S1519-69842006000200007>.
- Gonçalves-de-Freitas, E., Teresa, F.B., Gomes, F.S., Giaquinto, P.C., 2008. Effect of water renewal on dominance hierarchy of juvenile Nile tilapia. *Appl. Anim. Behav. Sci.* 112, 187–195. <http://dx.doi.org/10.1016/j.applanim.2007.07.002>.
- Good, I.J., Zellner, A., 1975. An introduction to bayesian inference in econometrics. *Technometrics* 17, 137. <http://dx.doi.org/10.2307/1268014>.
- Grobler, J.M.B., Wood, C.M., 2013. The physiology of rainbow trout in social hierarchies: two ways of looking at the same data. *J. Comp. Physiol. B Biochem. Syst. Environ. Physiol.* 183, 787–799. <http://dx.doi.org/10.1007/s00360-013-0752-5>.
- Haller, J., Wittenberger, C., 1988. Biochemical energetics of hierarchy formation in *Betta splendens*. *Physiol. Behav.* 43, 447–450. [http://dx.doi.org/10.1016/0031-9384\(88\)90118-7](http://dx.doi.org/10.1016/0031-9384(88)90118-7).
- Jaynes, E.T., 1976. Confidence intervals vs. Bayesian intervals. *Found. Probab. Theory, Stat. Inference, Stat. Theor. Sci.* [http://dx.doi.org/10.1007/978-94-010-1436-6\\_6](http://dx.doi.org/10.1007/978-94-010-1436-6_6).
- Jeffreys, H., 1961. *Theory of Probability*, 3rd ed. Oxford University Press, Oxford, UK.
- Kéry, M., Royle, J.A., 2016. Linear models, generalized linear models (GLMs), and random effects models. *Applied Hierarchical Modeling in Ecology*. Elsevier, pp. 79–122. <http://dx.doi.org/10.1016/B978-0-12-801378-6.00003-5>.
- Kinas, P.G., Andrade, H.A., 2007. Bayesian statistics for fishery stock assessment and management: a synthesis. *J. Aquat. Sci.* 2, 103–112.
- Kruschke, J., 2012. Bayesian estimation supersedes the t test. *J. Exp. Psychol. Gen.* <http://dx.doi.org/10.1037/a0029146>.
- Kruschke, J.K., 2015. *Metric predicted variable with one nominal metric predictor. Doing Bayesian Data Analysis: A Tutorial with R, JAGS and Stan*. Elsevier, pp. 553–581 (ISBN: 978-0-12-405888-0).
- Ludwig, D.A., 2005. Use and misuse of p-values in designed and observational studies: guide for researchers and reviewers. *Aviat. Space Environ. Med.* 76, 675–680.
- Maan, M.E., Groothuis, T.G.G., Wittenberg, J., 2001. Escalated fighting despite predictors of conflict outcome: solving the paradox in a South American cichlid fish. *Anim. Behav.* 62, 623–634. <http://dx.doi.org/10.1006/anbe.2001.1819>.
- McCarthy, M.A., 2007. *Introduction. Bayesian Methods for Ecology*. Cambridge University Press, pp. 1–29.
- McNamara, J.M., Green, R.F., Olsson, O., 2006. Bayes' theorem and its applications in animal behaviour. *Oikos* 112, 243–251. <http://dx.doi.org/10.1111/j.0030-1299.2006.14228.x>.
- Miyai, C.A., Carretero Sanches, F.H., Costa, T.M., Colpo, K.D., Volpato, G.L., Barreto, R.E., 2011. The correlation between subordinate fish eye colour and received attacks: a negative social feedback mechanism for the reduction of aggression during the formation of dominance hierarchies. *Zoology* 114, 335–339. <http://dx.doi.org/10.1016/j.zool.2011.07.001>.
- Montero, D., Lalumera, G., Izquierdo, M.S., Caballero, M.J., Saroglia, M., Tort, L., 2009. Establishment of dominance relationships in gilthead sea bream *Sparus aurata* juveniles during feeding: effects on feeding behaviour, feed utilization and fish health. *J. Fish Biol.* 74, 790–805. <http://dx.doi.org/10.1111/j.1095-8649.2008.02161.x>.
- Neat, F., Taylor, A., Huntingford, F., 1998. Proximate costs of fighting in male cichlid fish: the role of injuries and energy metabolism. *Anim. Behav.* 55, 875–882. <http://dx.doi.org/10.1006/anbe.1997.0668>.
- Pérez, E., Díaz, F., Espina, S., 2003. Thermoregulatory behavior and critical thermal limits of the angelfish *Pterophyllum scalare* (Lichtenstein) (Pisces: cichlidae). *J. Therm. Biol.* 28, 531–537. [http://dx.doi.org/10.1016/S0306-4565\(03\)00055-X](http://dx.doi.org/10.1016/S0306-4565(03)00055-X).
- Pinho-Neto, C.F., Miyai, C.A., Sanches, F.H.C., Giaquinto, P.C., Delicio, H.C., Barcellos, L.J.G., Volpato, G.L., Barreto, R.E., 2014. Does sex influence intraspecific aggression and dominance in Nile tilapia juveniles? *Behav. Processes* 105, 15–18. <http://dx.doi.org/10.1016/j.beproc.2014.02.003>.
- Pitak-Arnnop, P., Dhanuthai, K., Hemprich, A., Pausch, N.C., 2010. Misleading p-value: do you recognise it? *Eur. J. Dent.* 4, 356–358. <http://dx.doi.org/10.2460/jvma.241.6.690>.
- Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* 6, 7–11.
- R Development Core Team, 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria R Found. Stat. Comput. Vienna, Austria. <http://www.R-project.org/>.
- Raftery, A.E., 1995. Bayesian model selection in social research. *Sociol. Methodol.* <http://dx.doi.org/10.2307/271063>.
- Ros, A.F.H., Becker, K., Oliveira, R.F., 2006. Aggressive behaviour and energy metabolism in a cichlid fish, *Oreochromis mossambicus*. *Physiol. Behav.* 89, 164–170. <http://dx.doi.org/10.1016/j.physbeh.2006.05.043>.
- Sullivan, G.M., Feinn, R., 2012. Using effect size—or why the P value is not enough. *J. Grad. Med. Educ.* 4, 279–282. <http://dx.doi.org/10.4300/JGME-D-12-00156.1>.



- Terleph, T.A., 2004. The function of agonistic display behaviours in *Gnathonemus petersii*. *J. Fish Biol.* 64, 1373–1385. <http://dx.doi.org/10.1111/j.0022-1112.2004.00401.x>.
- Wagenmakers, E.-J., 2007. A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804. <http://dx.doi.org/10.3758/BF03194105>.
- Werner, M., Kramer, B., 2002. Intraspecific agonistic interactions in freely swimming mormyrid fish, *Marcusenius macrolepidotus* (South African form). *J. Ethol.* 20, 107–121. <http://dx.doi.org/10.1007/s10164-002-0062-y>.
- Wessel, M.L., Smoker, W.W., Fagen, R.M., Joyce, J., 2006. Variation of agonistic behavior among juvenile Chinook salmon (*Oncorhynchus tshawytscha*) of hatchery, hybrid, and wild origin. *Can. J. Fish. Aquat. Sci.* 63, 438–447. <http://dx.doi.org/10.1139/f05-227>.
- Wetzels, R., Grasman, R.P.P., Wagenmakers, E.-J., 2012. A default bayesian hypothesis test for ANOVA designs. *Am. Stat.* 66, 104–111. <http://dx.doi.org/10.1080/00031305.2012.695956>.
- Zellner, A., 1998. The finite sample properties of simultaneous equations' estimates and estimators Bayesian and non-Bayesian approaches. *J. Econ.* 6, 185–212.