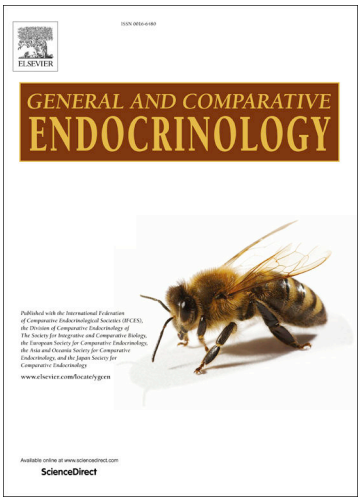
Journal Pre-proofs



Water-borne testosterone levels predict exploratory tendency in male poison frogs

Eva Ringler, Katharina Dellefont, Mélissa Peignier, Virginie Canoine

PII: S0016-6480(23)00221-6
DOI: https://doi.org/10.1016/j.ygcen.2023.114416
Reference: YGCEN 114416

To appear in: *General and Comparative Endocrinology*

Received Date: 27 June 2023
Revised Date: 22 October 2023
Accepted Date: 20 November 2023

Please cite this article as: Ringler, E., Dellefont, K., Peignier, M., Canoine, V., Water-borne testosterone levels predict exploratory tendency in male poison frogs, *General and Comparative Endocrinology* (2023), doi: https://doi.org/10.1016/j.ygcen.2023.114416

Water-borne testosterone levels predict exploratory tendency in male poison frogs

Eva Ringler[1,2,3,*], Katharina Dellefont[3,*], Mélissa Peignier[1,2,3], Virginie Canoine[3]

[1] Division of Behavioural Ecology, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland.

[2] Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University Vienna, University of Vienna, Vienna, Austria.

[3] Department of Behavioural and Cognitive Biology, University of Vienna, Vienna, Austria.

[*] shared first-authorship

Corresponding author:

Eva Ringler

Division of Behavioural Ecology, Institute of Ecology and Evolution

University of Bern

Wohlenstrasse 50a

3032 Hinterkapellen

SWITZERLAND

eva.ringler@unibe.ch

Abstract

Hormones play a fundamental role in mediating social behaviors of animals. However, it is less well understood to what extent behavioral variation between individuals can be attributed to variation in underlying hormonal profiles. The goal of the present study was to infer if individual androgen levels, and/or the modulation thereof, can explain among-individual variation in aggressiveness, boldness and exploration. We used as a model the dart-poison frog *Allobates femoralis* and took repeated non-invasive water-borne hormonal samples of individual males before (baseline) and after (experimental) a series of behavioral tests for assessing aggression, boldness, and exploratory tendency. Our results show that androgen levels in *A. femoralis* are quite stable across the reproductive season. Repeatability in $_{wb}$T baseline levels was high, while time of day, age of the frog, and trial order did not show any significant impact on measured $_{wb}$T levels. In general, experimental $_{wb}$T levels after behavioral tests were lower compared to the respective baseline levels. However, we identified two different patterns with regard to androgen modulation in response to behavioral testing: individuals with low baseline $_{wb}$T tended to have increased $_{wb}$T levels after the behavioral testing, while individuals with comparatively high baseline $_{wb}$T levels rather showed a decrease in hormonal levels after testing. Our results also suggest that baseline $_{wb}$T levels are linked to the personality trait exploration, and that androgen modulation is linked to boldness in *A. femoralis* males. These results show that differences in hormonal profiles and/or hormonal modulation in response to social challenges can indeed explain among-individual differences in behavioral traits.

Keywords: Testosterone, animal personality, behavioral challenge

1. Introduction

Several studies have demonstrated high within-individual consistency and between-individual variation of behavior across time and contexts in several animal taxa (i.e. termed *animal personality*; Araya-Ajoy and Dingemanse, 2014; Réale et al., 2007). Empirical and theoretical approaches have demonstrated how such differences in behavior ultimately affect an animal's prospects of survival, competitive ability, mating success and other fitness relevant traits (Dingemanse et al., 2004; Sih and Bell, 2008). One key question in the study of animal personality is to what extent differences in behavioral phenotypes can be attributed to constraints imposed by underlying physiology, such as hormonal profiles (Groothuis and Carere, 2005). In turn, behavioral phenotypes might also arise from differences across individuals in their physiological response to challenges in their social/natural environment (e.g. Biro and Stamps, 2010; Fürtbauer et al., 2015; Réale et al., 2010; Sih et al., 2015).

Hormones play a fundamental role in the expression of morphological and behavioral traits (Adkins-Regan, 2005). The causal relationship between hormones and behavior is bidirectional, as hormones regulate the expression of social behaviors, but at the same time being exposed to behavior of con- or heterospecifics can also induce a hormonal response in the focal individual (Adkins-Regan, 2005; Gabor and Grober, 2010; Vitousek et al., 2014; Wingfield et al., 1990). The precise interplay between hormones and behavior may differ between individuals of a population, as it depends on various factors, such as experience during early development, an animal's physiological condition, environment and adaptation to specific life-history stages – all of which may lead to a variation in behavioral phenotypes.

Animal personality is typically measured along five main axes: aggressive-docile, exploration affine-averse, sociable-solitary, bold-shy and active-passive (Réale et al., 2007). Several behaviors across these five axes are known to be modulated by sex steroids, precisely androgens and especially during the

reproductive season (Hau, 2007; Nelson, 2005). In male vertebrates, androgens play a key role in the development and maintenance of primary and secondary sexual traits but also regarding the modulation of different behaviors related to reproduction, such as courtship, mating behavior or territorial aggression (Burmeister and Wilczynski, 2001; Fusani, 2008; Hirschenhauser et al., 2003; Hunt et al., 2019; Rosvall et al., 2020; Rosvall et al., 2012). However, persistent high levels of testosterone (T) are expected to bear considerable costs (Wingfield et al., 2001), such as reduced immune function (Dufty, 1989; McGlothlin and Ketterson, 2008), increased risk-taking and resulting elevated predation risk (Marler and Moore, 1988; Raynaud and Schradin, 2014), interference with parental and other social behavior (Fürtbauer et al., 2020; Hegner and Wingfield, 1987; McGlothlin et al., 2007), and metabolic expenses (Buchanan et al., 2001; Tobler et al., 2007). To minimize these negative impacts, androgens undergo seasonal fluctuations, with the highest concentration during the breeding season and lowest during the non-reproductive period (Goymann et al., 2019; Hau, 2007; Husak et al., 2021). High among- and within-species differences in the levels of circulating T have been shown for many vertebrate taxa, and these differences have been linked to respective social and environmental factors, such as breeding season length, type of mating system, and latitude (cf. Husak et al., 2021). Also, on smaller timescales, fine-tuned temporal fluctuations in T modulating behavioral and physiological responses to sudden environmental challenges may reflect the trade-off associated with high and low levels of circulating T in males (Hunt et al., 1995; Rodríguez et al., 2022; Romero et al., 1998; Wingfield et al., 1990). Identifying the ecological and physiological factors that shape behavior at the species, population, but also individual level will advance our understanding about the mechanisms that underline behavioral variation across these different levels of biological organization. Previous research identified a prominent role of glucocorticoids (e.g. cortisol) for explaining differences in behavioral profiles (Baugh et al., 2017; Baugh et al., 2012; Dosmann et al., 2015; Grace and Anderson, 2014), due to their prominent role in stress-axis-programming. Only few studies have looked how androgens shape individual personalities (but see Hau and Goymann, 2015; Kraus et al., 2020; Mutzel et al., 2011).

The aim of the present study was to infer if individual baseline androgen levels are repeatable (i.e. if hormone levels show low within-individual and high between-individual variation) and can explain among-individual variation in the personality traits aggressiveness, boldness and exploration. Furthermore, we asked if also the modulation of androgen levels after a suite of behavioral challenges is repeatable across individuals. We used as a model the Neotropical poison frog *Allobates femoralis,* a territorial species that shows highly aggressive behavioral response towards acoustic playbacks, simulating calling intruders (Hödl, 1983; Narins et al., 2003; Ringler et al., 2011; Rodríguez et al., 2022). We took repeated samples of individual 'baseline' water borne testosterone ($_{wb}$T) levels (i.e., without any prior behavioral manipulation) to assess repeatability in individual hormonal profiles over the course of the study period. Furthermore, we took hormonal samples directly after a series of behavioral tests (hereafter 'experimental' $_{wb}$T), which assessed territorial aggression, boldness and explorative behavior of individual males. For the hormonal sampling, we used a non-invasive water bath method (Baugh et al., 2018; Gabor et al., 2013; Rodríguez et al., 2022), which enabled us to take repeated samples of the same individuals in the field. We assessed among- and within-individual consistency of $_{wb}$T to gain information about the consistency of individual hormonal profiles over the course of several weeks during the breeding season. Additionally, we identified if and how preceding behavioral testing will alter obtained hormonal measurements. The combination of repeated hormonal sampling and behavioral testing allowed us to link the endocrine profile and modulation of each individual to its personality.

We expected $_{wb}$T levels to be highly repeatable, especially when sampled without any preceding behavioral manipulation. We also expected $_{wb}$T to be positively linked to levels of territorial aggression, boldness, and/or exploration in the behavioral tests, given that previous studies in several species, including *A. femoralis*, suggested a link between androgen modulation and spatial behavior in the context of territory defense or homing (cf. Herman and Wallen, 2007; Hodgson et al., 2008; Pašukonis et al., 2022; Rodríguez et al., 2022).

2. Materials and methods

2.1 Study site and study species

This study was conducted in a free-ranging population of *A. femoralis* on a river island of approx. 5ha, close to the field camp 'Saut Pararé' (4°02′ N, 52°41′ W) in the nature reserve 'Les Nouragues', in French Guiana (Bongers et al., 2001; Ringler et al., 2016). The island population of *A. femoralis* was established in 2012 by introducing tadpoles from the nearby mainland population and has been stable ever since with approximately 150 individuals (Ringler et al., 2015). We conducted fieldwork during the rainy season, from the beginning of February 2019 until the end of April 2019, which coincides with the reproductive season of the focal species (Gottsberger and Gruber, 2004).

Allobates femoralis (Boulenger 1883) is a small, diurnal Neotropical poison frog (Dendrobatidae *sensu* AmphibiaWeb, 2023), which is distributed throughout the Amazon Basin and Guiana Shield. During the reproductive season, males are highly territorial and advertise territory occupancy to male competitors and potential female mating partners by producing loud advertisement calls from exposed, elevated positions (Hödl, 1983; Ringler et al., 2011; Rodríguez López and Hödl, 2020; Roithmair, 1992). Males vigorously defend their territory against conspecific intruders (Narins et al., 2003). Females exhibit site fidelity but are typically not aggressive towards either sex (Ringler et al., 2012; Ringler et al., 2009), and actively approach neighboring calling males for courtship and mating (Montanarin et al., 2011; Stückler et al., 2019). Egg deposition takes place in the male's territory and both sexes mate multiple times with multiple partners (Ringler et al., 2012; Ursprung et al., 2011). After hatching, tadpoles are typically transported by the male to medium sized water bodies located up to 200m outside the territory (Beck et al., 2017; Ringler et al., 2018; Ringler et al., 2013).

2.2 Population monitoring

We surveyed the entire population every day from 0900 to 1800 h. We identified all frogs on site via digital pictures of their unique ventral patterns and later verified their identity with the pattern matching software Wild-ID (Bolger et al., 2012). Frogs were sexed by the presence (males) or absence (females) of a vocal sac. We recorded the precise location of the frogs on a digital map, using a tablet PC (WinTab 9, Odys, Willich, Germany) equipped with the mobile GIS software ArcPad 10.2 (ESRI, Redlands, CA, U.S.A.). We determined body size (snout urostyle length) from dorsal photographs taken on top of a measurement grid using the software Image J 1.52a (Rasband, 1997-2021). Information on the age of individuals was available from a concurrent long-term monitoring on the island population since its origin in 2012.

2.3 Experimental design

To gain information about the among- and within-individual variation of individual $_{wb}$T levels and further investigate the effect of preceding behavioral tests on their T response, we repeatedly sampled $_{wb}$T under two following conditions: First we collected 'baseline' levels by capturing a focal frog without the use of any acoustic stimuli (e.g. playback) and immediately transferred it to the water bath (for details see '*Hormonal sampling*' and Figure 1). Second, we also measured 'experimental' $_{wb}$T levels immediately after the focal individual had completed a consecutively deployed suite of behavioral tests to assess personality traits (for details see '*Behavioral experiments*' and Figure 1). In every trial we noted the date and time of

day (am or pm) when the measurement was taken, as well as individual parameters such as body size (in mm) and age, measured as a binomial variable (first reproducer vs. recapture from previous years). We aimed for obtaining three replicates in each condition per individual, summing up to a total of six measurements per frog. Half of the tested individuals started with 'baseline', while the other half started with the 'experimental' sampling. Consecutive samples were always taken more than 24 hours apart. After every second trial we added a break of at least 3 days to minimize any confounding effects of the procedure on the measurements.

2.4 Behavioral tests

All individuals underwent a set of behavioral tests to quantify the following behavioral traits: territorial aggression, boldness and explorative tendency. The procedure of these combined tests lasted for a total of about 30 min and to facilitate reading, we will from now on define both tests with 'behavioral test' unless we specifically refer to one of these tests only.

Territorial aggression: We assessed within- and between-individual variation in the levels of territorial aggression in individual males by simulating a calling intruder inside a male's territory. To do so, we used a simulated territorial intrusion test to induced territorial defense behavior of the territorial male by broadcasting /presenting synthetic conspecific call by a loudspeaker (for details see Peignier et al., 2022). These conspecific male calls elicit aggressive responses of a territorial male (Rodríguez et al., 2022; Sonnleitner et al., 2020; Ursprung et al., 2009) which can be categorized/quantified in following behavioral parameters: a) latency until the first head-body orientation and b) until the first jump, c) the likelihood to jump in moments when the speaker was silent (i.e., between bouts of calls), and d) the speed to approach the speaker (cf. Chaloupka et al., 2022; Peignier et al., 2022).

Boldness and Exploration: Immediately after the previous test, we caught the frog and assessed exploration- and boldness-related behaviors using a Novel Environment Setup (cf. Peignier et al., 2022). The setup consisted of a cooler box (hereafter "Novel Environment"), with a PVC tube attached on one side of the box (hereafter "shelter"). We first put the frog in the dark shelter for five minutes, to allow the individual to acclimatize to the setup. Afterwards we opened the shelter and allowed the focal frog to explore the Novel Environment for 15 minutes. We measured the a) latency and b) probability to leave the shelter as well as c) the distance travelled, d) the number of jumps performed, and e) the area covered in the novel environment (for more details see Peignier et al., 2022).

To assess within- and between-individual variation in behavior we repeated those tests several times: we conducted 163 territorial defense tests with 51 males (mean ± SD = 3.20 ± 1.31 repetitions per individual) and 156 Novel Environment Tests with 50 males (mean ± SD = 3.31 ± 1.50 repetitions per individual). In a previous study, using the same behavioral dataset, it was shown that the behaviors measured during the both tests are repeatable and that the latency until the first jump, the distance travelled, and the time spent in the shelter best represented aggression, exploration, and boldness, respectively (Peignier et al., 2022). In the present study, we use these measures as proxies for the personality traits aggression, exploration and boldness to investigate the link between personality and $_{wb}$T levels.

2.5 Hormonal sampling and analysis

We used a non-invasive water-bath method (Baugh et al., 2018; Baugh and Gray-Gaillard, 2021; Gabor et al., 2013) with variations following the protocol described in Rodríguez et al. (2022) to collect repeated $_{wb}$T

measurements of male *A. femoralis*. In brief, after capture, we put the frogs in a small glass box (14cm x 9cm x 5cm), filled with 40 mL of distilled water and left them in this box for one hour (Figure 1). Resulting concentrations (pg/mL) thus represent androgen release rates of one individual over one hour. The dimensions of the box and the water volume were chosen so that the frogs' body was covered with water, water levels did not constrain breathing (nose was outside water), and frogs were not able to climb out of the water. An opaque cover was placed over the box to minimize any disturbances from outside and to minimize stress of the focal individual. After one hour the frog was gently released at the original capture location. Non-polar hormones were extracted by processing each water sample through 20 mL sterile syringes coupled to an individual C18 cartridge (SPE, Sep-Pak C18 Plus, 360 mg Sorbent, 55–105 µm particle size, #WAT020515, Waters corp., Milford, MA) with a flow rate of ca. 10 mL/min. Afterwards, cartridges were eluted with 4 mL of 96% EtOH into 8 mL borosilicate vials and stored at 4 °C until further processing in the endocrinological lab at the University of Vienna. Between water-bath samplings, water bowls were thoroughly rinsed with ethanol and distilled water, and fully dried before subsequent use. Researchers were wearing nitrile gloves at all times, which were changed between each sample, to avoid contamination.

In order to quantify $_{wb}$T (in pg/mL), we used a commercially available ELISA kit (Enzo Life science #ADI-900 065). Beforehand, 1mL out of the 4mL of 96% EtOH eluded samples were pipetted into a glass tube and dried down under a N_2-stream and then re-suspended in 250ul Assay buffer provided by the manufacturer. Preliminary tests have shown that 1 mL of 96% EtOH was sufficient to quantify reliably testosterone concentration. Because the antibody has a very low cross-reaction with other androgens (19-hydroxytestosterone <15%, androstendione <7.2%, Estradiol < 0.4, all others < 0,001%) we dare to assume to have mostly measured testosterone. Final concentration of the samples was corrected for dilution factor. The detection limit for the assay was 5.67 pg/mL. The intra-assay CV% of all duplicates was below 5.3%. The inter-assay CV% was calculated using a control sample and was below 11% (*n=8*).

2.6 Statistical analysis

The statistical analyses were conducted in RStudio (RStudio Team, 2019). We log transformed the $_{wb}$T measurements as it deviated from normality. Where possible, we report results as $p > 0.1$ no evidence, $0.1 < p < 0.05$ weak evidence, $0.05 < p < 0.01$ moderate evidence, $0.01 < p < 0.001$ strong evidence, $p < 0.001$ very strong evidence (Muff et al., 2022).

To investigate factors that affect the overall androgen levels, we fitted a linear mixed model using the function 'lmer', in the package 'lme4', with condition ('baseline'/'experimental'), time of day (am/pm), age (new encounters/survivors from a previous reproductive season), body size, whether the frog was calling or not before the hormonal measurement, and trial order as fixed effects. We included ID as random effect, and $_{wb}$T concentration (log transformed) as response variable. We assured that model assumptions of residual normality were met by visually inspecting qq-plots. The condition was the only factor influencing overall $_{wb}$T level. We further investigated the consistency of $_{wb}$T levels within and between individuals, using both reduced ('baseline' only or 'experimental' only) datasets, with the 'rpt' function in the rptR package (Stoffel et al., 2017). To identify if $_{wb}$T levels at 'baseline' itself had an influence on androgen modulation during/after the behavioral tests, we calculated Δ_{wb}T by subtracting individual average 'baseline' levels from the respective average 'experimental' levels from all individual males. We then tested for a possible correlation between 'baseline' $_{wb}$T and Δ_{wb}T using a Pearson correlation test.

We also studied how aggression, exploration and boldness covary with $_{wb}$T levels and modulation at the among- and within-individual level using a Markov chain Monte Carlo method. We built two Bayesian linear mixed effect models (Hadfield, 2010) with the three personality scores as response variables

and ID as random factor. In addition, we added as response variables the log transformed 'baseline' $_{wb}T$ in the first model and the $\Delta_{wb}T$ in the second model. We scaled each response variable by centering to their mean value and standardizing to units of 1 phenotypic standard deviation. We estimated the among- and within-individual correlations and covariances between each of the personality score and the $_{wb}T$ using the posterior distributions. We used an uninformative prior and ran 2,000,000 iterations with a burn-in of 80,000, and selected every 750th posterior parameter sample after the initial burn-in. We assumed statistical significance if the 95% credible intervals did not overlap 0. We assured that model assumptions were met by verifying the absence of autocorrelation (correlation between lags <0.1; Hadfield, 2010), sufficient mixing (plots of Markov-Chain-Monte-Carlo chains), and performing a Heidelberg and Welch diagnostic test.

3. Results

In total we collected 252 hormonal samples from 40 individual males. We obtained samples for baseline $_{wb}T$ from 37 males ('baseline': 1–6 samples per male, mean ± SD = 3.51 ± 1.19 samples per male), and samples after the behavioral manipulations from 39 males ('experimental': 1–5 samples per male; mean ± SD = 3.13 ± 1.08 samples per male).

We did not find any evidence that hormonal measurements were influenced by the time of day when samples were collected, the age or body size of the individual, the activity (i.e., calling or not) of the male, or the trial order (all $p > 0.05$; Table 1). However, we found very strong evidence that $_{wb}T$ levels were lower ($\beta = -0.3$, $p < 0.001$, Table 1) when measured after behavioral experiments (mean average $_{wb}T$ ± SD = 306.15 ± 107.5 pg/mL) compared to the respective baseline samples (mean average $_{wb}T$ ± SD = 387.43 ± 171.1 pg/mL). Repeatability was quite high for the baseline samples ('baseline': R = 0.45; 95%CI = [0.24;0.61]), but much lower when hormones were collected after behavioral manipulations ('experimental': R = 0.24; 95%CI = [0.03;0.43]).

Not all individuals responded to the behavioral manipulations with a reduction in $_{wb}T$. Interestingly, we found a very strong evidence for a negative correlation between average baseline $_{wb}T$ and $\Delta_{wb}T$ in males (Pearson correlation test: $r = -0.76$, $t = -6.81$, $df = 33$, $p < 0.001$, Figure 2A). This means, that individuals with a relatively low baseline $_{wb}T$ tended to increase hormone levels after the behavioral tests, while individuals with a comparatively high baseline level showed a decrease in their androgen levels after the behavioral manipulation (Figure 2B).

Our results show a clear trend for a positive covariation between exploration and baseline $_{wb}T$ levels at the among- and at the within-individual level, as confidence intervals only slightly overlapped zero (among-individual level: estimate = 0.18, 95% CI = [-0.01, 0.45]; within-individual level: estimate = -0.14, 95% CI = [-0.31, 0.03]; Table 2). Males showed increased levels of exploration when their baseline $_{wb}T$ levels were high. We also found a clear trend for a covariation between boldness and androgen modulation during/after the behavioral tests at the within-individual level (estimate = -0.18, 95% CI = [-0.39, 0.03]; Table 2). Males that showed the highest increase in $_{wb}T$ levels after the behavioral tests were also very bold.

4. Discussion

4.1 Repeatability in androgen levels

Our results show that male *A. femoralis* have relatively consistent $_{wb}$T levels throughout the breeding season. The repeatability of hormonal measurements that were obtained from non-invasive water bath samples was quite high, especially for the 'baseline' samples (R = 0.45). These values were considerably higher than repeatability scores of glucocorticoid hormones across several vertebrates (Schoenemann and Bonier, 2018). This indicates that T levels are relatively constant over the course of several weeks within the reproductive season in *A. femoralis* and that there are consistent differences between individuals in their baseline androgen levels. As a consequence, even a low number of repeated measurements allowed reliable estimations of individual baseline hormonal profiles in male poison frogs.

Further, we did not find a significant difference between hormonal samples collected in the morning and in the afternoon. This was contrary to what we expected, as in a previous study androgen levels were found to be increased in the afternoon compared to morning hours, which was linked to general calling activity in a nearby *A. femoralis* population (Rodríguez et al., 2022). In vertebrates, steroid concentrations commonly undergo a circadian rhythm; they increase during early morning hours and drop in the afternoon (Nelson, 2005). Several studies have documented the existence of diurnal cycles of circulating T (fish: Lorenzi et al., 2008; monkeys: Schlatt et al., 2008; humans: Diver et al., 2003; but see also Licht et al., 1985 for green sea turtles). In several animal taxa, T concentrations are positively related to latitude and negatively to the length of the breeding season (Eikenaar et al., 2012; Husak et al., 2021), which suggests that tropical animals usually exhibit lower T levels with very low seasonal fluctuation during the reproductive season compared to temperate-zone species (see also Canoine et al., 2007; Hau et al., 2008). However, it is possible that other environmental factors (e.g. predators, temperature, climatic conditions, calling activity), might have contributed to the differences in diurnal T variation found in this and the study of Rodríguez et al., 2022.

The factor 'age' did not show a significant relationship with androgen levels. There is evidence for an age-related change of T levels in various animal taxa (Groothuis and Carere, 2005; Schlatt et al., 2008; Těšický et al., 2022), however those taxa typically show a greater lifespan than our studied species. In *A. femoralis* the majority of the population only survives one reproductive season (cf. Ringler et al., 2015), and therefore age likely is not a relevant factor for the variation in androgen levels in this short-lived species.

4.2 Influence of behavioral tests on androgen levels

An increase in T has been observed in many species following social challenges (Goymann et al., 2019; Wingfield et al., 2020; see also Assis et al., 2012; Leary, 2014; Moore et al., 2020). In our study, $_{wb}$T levels generally dropped after the behavioral tests. This is in contrast with a recent study which found a positive androgenic response to simulated territorial intrusions in *A. femoralis* males, providing support for the Challenge Hypothesis (Rodríguez et al., 2022). In this previous study, water-borne androgen levels were elevated after presenting a conspecific playback compared to baseline conditions, but only in males that actually approached the loudspeaker and not in males which did not react to the playback. Moreover, while an increase of $_{wb}$T was observed within the first hour of water sampling after playback presentation, androgen levels clearly dropped in the following hours, even below baseline levels. Curiously, the 'non-responders' showed a much stronger decline of $_{wb}$T three hours after the playback test compared to males who actively approached the loudspeaker. In this previous study, the decline of $_{wb}$T levels was observed only after more than 2 hours following exposure to a behavioral test/or stressor, and might have been due to the activation of the negative feedback system of the hypothalamic–pituitary–gonadal (HPG) axis, or due to other hormones, such as glucocorticoids, via an antagonistic effect on the HPG axis (Moore and Jessop 2003; see also "Energetics-Hormone Vocalization hypothesis" by Emerson and Hess 2001). We cannot fully exclude similar effects may have impacted on our measurements, but since all hormonal samples were collected in the same standardized procedure, we assume that such effect might have impacted all samples

equally. Because in the present study we collected the hormonal samples not immediately after the territorial aggression test, but after a suite of behavioral tests which lasted in total about 30 min, the resulting androgen levels actually represent a combined/integrated hormonal response to the entire test sequence. Further studies are needed to clarify the link between the hypothalamic–pituitary–adrenal (HPA) and the HPG axis in *A. femoralis*.

Most interestingly, not all individuals responded to the behavioral tests in the same way. We observed two different patterns when comparing 'baseline' and 'experimental' $_{wb}$T levels: individuals with low baseline $_{wb}$T tended to show an increase in hormonal levels, while individuals with comparatively high baseline $_{wb}$T rather showed a decrease in hormonal levels following the behavioral manipulation (Figure 2). This phenomenon could happen if the physiological maximum of individuals' high baseline $_{wb}$T was already reached before the start of the behavioral tests, e.g. due to a stressful interaction. These individuals might not be able to further increase their androgen levels when faced with a new challenge, while individuals with low baseline $_{wb}$T could (Goymann et al., 2007; Wingfield et al., 1990). However, because the repeatability of baseline $_{wb}$T was so high, it is unlikely that these individuals were caught each time just prior a stressful event. Alternatively, our results could potentially be caused by differential personality types being linked to differential physiological (i.e. hormonal) responses to stress and/or social challenges. Previous studies have shown a link between animal personality and differential physiological response to social challenges (i.e. "stress coping styles"; Baugh et al., 2017; Baugh et al., 2012; Carere and Maestripieri, 2013; Cockrem, 2013; Coppens et al., 2010). For example, proactive and reactive rats and mice differed in the levels of serotonin release or receptor expression in the prefrontal cortex (Caramaschi et al., 2007; Ferrari et al., 2003; van Erp and Miczek, 2000). Although we did not find a link between hormonal modulation and any personality trait we measured, we cannot rule out that another underlying trait, such as activity, is responsible for the observed pattern.

4.3 Link between personality traits and androgen levels and modulation

Interestingly, we found no evidence that 'baseline' androgen levels were associated with territorial aggression in *A. femoralis* males. Similar results were reported in a previous study where the intensity of the phonotactic approach towards a playback was not related to androgen responsiveness in *A. femoralis* males (Rodríguez et al., 2022). However, our results suggest that $_{wb}$T levels are positively linked to the personality trait 'exploration'. Individuals with a high exploratory tendency in the Novel Environment Test also had high baseline $_{wb}$T levels. These findings are in line with a recent study that found androgens to be associated with navigation-associated behaviors in three species of poison frogs (Pašukonis et al., 2022). Higher baseline androgen levels were found in individuals that also showed more exploration after translocation in *D. tinctorius*. The amount of exploration during the navigation task was associated with an increase in androgen levels in *A. femoralis*, while successful homing was found to result in a significant decrease in androgen compared to baseline levels (Pašukonis et al., 2022). These results together with the findings of our present study highlight a prominent role of androgens for among-individual variation in exploration-related spatial behaviors in males.

Our results further suggest a link between T modulation and boldness. Males that showed the highest increase in $_{wb}$T levels after the behavioral tests were also very bold. These findings are in line with the Challenge Hypothesis, which states that the temporary increase of T is to facilitate potentially challenging or threatening encounters (Wingfield et al., 1990). As our experimental design does not allow to disentangle cause and consequence of these correlations, future studies using hormonal manipulation experiments in combination with behavioral experiments are needed to precisely disentangle the causal relationship between T modulation and exploration behavior and boldness.

373

374 *4.4 Methodological implications*

375 One key question in animal personality research is why there are consistent individual differences in
376 behavior. One hypothesis is that heritable traits leading to genetically determined physiological differences
377 among individuals could give rise to consistent behavioral differences (Baugh et al., 2012; Drent et al.,
378 2003; Stamps and Groothuis, 2010). The present study highlights the importance of the exact time point
379 when the hormonal sample is collected, for trying to link hormonal with behavioral profiles. To minimize
380 handling time and experimental effort, it would be ideal to measure hormonal levels directly before and/or
381 after behavioral tests, however this procedure could affect the hormonal and behavioral experiments,
382 respectively. Whether hormonal measurements should be taken independently from behavioral tests or
383 immediately thereafter, ultimately depends on the research question. To gain information about long-term
384 differences in hormonal profiles, measurements should be taken without any preceding behavioral
385 experiments. However, when determining differences in physiological responses after behavioral
386 challenges, it is important to carry out hormonal sampling completely independently from other
387 experimental manipulations as well as directly after behavioral testing.

388

389 *4.5 Conclusions*

390 Our findings show that androgen levels in *A. femoralis* are quite stable across the reproductive season, as
391 repeatability in $_{wb}$T baseline levels was high. Time of day, age and calling activity had no influence on
392 hormone levels. As expected, preceding behavioral experiments had a strong influence on the variation in
393 $_{wb}$T levels, but the direction of the hormonal response appears to be linked to individual baseline levels.
394 Individuals with low baseline levels showed an increase in androgen, while individuals with high baseline
395 levels showed a reduction of androgen levels after the behavioral tests. In addition, we show that *in A.*
396 *femoralis*, androgens are linked to the personality trait exploration. Taken together, these results suggest
397 that differences in hormonal profiles and/or responses to social challenges can – at least partially – explain
398 among-individual differences in behavioral traits.

399

400 **Ethics**

401 This study was approved by the scientific committee of the 'Nouragues Ecological Research Station' and
402 the ethics and animal welfare committee of the University of Veterinary Medicine Vienna. The hormonal
403 such as the behavioral sampling was conducted in strict accordance with current French and EU law,
404 according to the Study of Animal Behaviour (ASAB) guidelines.

405

406 **Funding**

407 This study was funded by the Austrian Science Fund (FWF) via the project P 31518 (PI: Eva Ringler).

408

409 **CRediT authorship contribution statement**

VC and ER designed the study. KD and MP collected the data. MP performed the statistical analysis and did the data curation. KD wrote the first draft of the manuscript. MP, VC and ER reviewed, edited and wrote sections of the manuscript. VC and ER supervised the study. All authors contributed to manuscript revision, and have read, and approved the submitted version.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

We thank the Nouragues research field station (managed by CNRS), which benefits from 'Investissement d'Avenir' grants managed by the Agence Nationale de la Recherche (AnaEE France ANR-11-INBS-0001; Labex CEBA ANR-10-LABX-25-01). We are grateful to the staff of CNRS Guyane for logistic support in French Guiana. Many thanks to Camilo Rodriguez Lopez for providing help and advice regarding the hormonal sampling, to Sarah Chaloupka and Christoph Leeb for help with fieldwork, and to Birgit Szabo for valuable comments on the manuscript.

Data availability

The datasets generated during and/or analyzed during the current study are available in the Open Science Framework repository: https://osf.io/67wvj/?view_only=48daf77111894781bd89af222c26be3e

References

Adkins-Regan, E., 2005. Hormones and animal social behavior. Princeton University Press.

AmphibiaWeb, 2023. AmphibiaWeb: Information on amphibian biology and conservation. University of California. https://amphibiaweb.org (accessed 16 March 2023).

Araya-Ajoy, Y.G., Dingemanse, N.J., 2014. Characterizing behavioural 'characters': an evolutionary framework. Proc R Soc Lond B 281, 20132645. https://doi.org/10.1098/rspb.2013.2645.

Assis, V.R. de, Navas, C.A., Mendonça, M.T., Gomes, F.R., 2012. Vocal and territorial behavior in the Smith frog (*Hypsiboas faber*): relationships with plasma levels of corticosterone and testosterone. Comp Biochem Physiol A Mol Integr 163, 265–271. https://doi.org/10.1016/j.cbpa.2012.08.002.

Baugh, A.T., Bastien, B., Still, M.B., Stowell, N., 2018. Validation of water-borne steroid hormones in a tropical frog (*Physalaemus pustulosus*). Gen Comp Endocrin 261, 67–80. https://doi.org/10.1016/j.ygcen.2018.01.025.

Baugh, A.T., Gray-Gaillard, S.L., 2021. Excreted testosterone and male sexual proceptivity: A hormone validation and proof-of-concept experiment in túngara frogs. Gen Comp Endocrin 300, 113638. https://doi.org/10.1016/j.ygcen.2020.113638.

Baugh, A.T., Schaper, S.V., Hau, M., Cockrem, J.F., Goede, P. de, van Oers, K., 2012. Corticosterone responses differ between lines of great tits (*Parus major*) selected for divergent personalities. Gen Comp Endocrin 175, 488–494. https://doi.org/10.1016/j.ygcen.2011.12.012.

Baugh, A.T., Senft, R.A., Firke, M., Lauder, A., Schroeder, J., Meddle, S.L., van Oers, K., Hau, M., 2017. Risk-averse personalities have a systemically potentiated neuroendocrine stress axis: A multilevel experiment in *Parus major*. Horm Behav 93, 99–108. https://doi.org/10.1016/j.yhbeh.2017.05.011.

Beck, K.B., Loretto, M.-C., Ringler, M., Hödl, W., Pašukonis, A., 2017. Relying on known or exploring for new? Movement patterns and reproductive resource use in a tadpole-transporting frog. PeerJ 5, e3745. https://doi.org/10.7717/peerj.3745.

Biro, P.A., Stamps, J.A., 2010. Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? Trends Ecol Evol 25, 653–659. https://doi.org/10.1016/j.tree.2010.08.003.

Bolger, D.T., Morrison, T.A., Vance, B., Lee, D., Farid, H., 2012. A computer-assisted system for photographic mark–recapture analysis. Methods Ecol Evol 3, 813–822. https://doi.org/10.1111/j.2041-210X.2012.00212.x.

Bongers, F., Charles-Dominique, P., Forget, P.-M., Théry, M. (Eds.), 2001. Nouragues: Dynamics and plant-animal interactions in a Neotropical rainforest. Kluwer Academic Publishers, Dordrecht, Netherlands.

Buchanan, K.L., Evans, M.R., Goldsmith, A.R., Bryant, D.M., Rowe, L.V., 2001. Testosterone influences basal metabolic rate in male house sparrows: a new cost of dominance signalling? Proc R Soc Lond B 268, 1337–1344. https://doi.org/10.1098/rspb.2001.1669.

Burmeister, S.S., Wilczynski, W., 2001. Social context influences androgenic effects on calling in the green treefrog (*Hyla cinerea*). Horm Behav 40, 550–558. https://doi.org/10.1006/hbeh.2001.1723

Canoine, V., Fusani, L., Schlinger, B., Hau, M., 2007. Low sex steroids, high steroid receptors: Increasing the sensitivity of the nonreproductive brain. Dev Neurobiol 67, 57–67. https://doi.org/10.1002/dneu.20296.

Caramaschi, D., Boer, S.F. de, Koolhaas, J.M., 2007. Differential role of the 5-HT1A receptor in aggressive and non-aggressive mice: an across-strain comparison. Physiol Behav 90, 590–601. https://doi.org/10.1016/j.physbeh.2006.11.010.

Carere, C., Maestripieri, D. (Eds.), 2013. Animal personalities: Behavior, physiology, and evolution. University of Chicago Press, Chicago, IL, USA.

Chaloupka, S., Peignier, M., Stückler, S., Araya-Ajoy, Y., Walsh, P., Ringler, M., Ringler, E., 2022. Repeatable territorial aggression in a Neotropical poison frog. Front Ecol Evol 10, 881387. https://doi.org/10.3389/fevo.2022.881387.

Cockrem, J.F., 2013. Individual variation in glucocorticoid stress responses in animals. Gen Comp Endocrin 181, 45–58. https://doi.org/10.1016/j.ygcen.2012.11.025.

Coppens, C.M., Boer, S.F. de, Koolhaas, J.M., 2010. Coping styles and behavioural flexibility: towards underlying mechanisms. Phil Trans R Soc Lond B 365, 4021–4028. https://doi.org/10.1098/rstb.2010.0217.

Dingemanse, N.J., Both, C., Drent, P.J., Tinbergen, J.M., 2004. Fitness consequences of avian personalities in a fluctuating environment. Proc R Soc Lond B 271, 847. https://doi.org/10.1098/rspb.2004.2680.

Diver, M.J., Imtiaz, K.E., Ahmad, A.M., Vora, J.P., Fraser, W.D., 2003. Diurnal rhythms of serum total, free and bioavailable testosterone and of SHBG in middle-aged men compared with those in young men. Clin endocrinol 58, 710–717. https://doi.org/10.1046/j.1365-2265.2003.01772.x.

Dosmann, A.J., Brooks, K.C., Mateo, J.M., 2015. Within-individual correlations reveal link between a behavioral syndrome, condition and cortisol in free-ranging Belding's ground squirrels. Ethology 121, 125–134. https://doi.org/10.1111/eth.12320.

Drent, P.J., van Oers, K., van Noordwijk, A.J., 2003. Realized heritability of personalities in the great tit (*Parus major*). Proc R Soc Lond B 270, 45–51. https://doi.org/10.1098/rspb.2002.2168.

Dufty, A.M., 1989. Testosterone and survival: A cost of aggressiveness? Horm Behav 23, 185–193. https://doi.org/10.1016/0018-506x(89)90059-7.

Eikenaar, C., Husak, J., Escallón, C., Moore, I.T., 2012. Variation in testosterone and corticosterone in amphibians and reptiles: relationships with latitude, elevation, and breeding season length. Am Nat 180, 642–654. https://doi.org/10.1086/667891.

Emerson, S.B., Hess, D.L., 2001. Glucocorticoids, androgens, testis mass, and the energetics of vocalization in breeding male frogs. Horm. Behav. 39, 59–69. https://doi.org/10.1006/hbeh.2000.1635.

Ferrari, P.F., van Erp, A.M.M., Tornatzky, W., Miczek, K.A., 2003. Accumbal dopamine and serotonin in anticipation of the next aggressive episode in rats. Eur J Neurosci 17, 371–378. https://doi.org/10.1046/j.1460-9568.2003.02447.x.

Fürtbauer, I., Brown, M.R., Heistermann, M., 2020. Collective action reduces androgen responsiveness with implications for shoaling dynamics in stickleback fish. Horm Behav 119, 104636. https://doi.org/10.1016/j.yhbeh.2019.104636.

Fürtbauer, I., Pond, A., Heistermann, M., King, A.J., 2015. Personality, plasticity and predation: linking endocrine and behavioural reaction norms in stickleback fish. Funct Ecol 29, 931–940. https://doi.org/10.1111/1365-2435.12400.

Fusani, L., 2008. Testosterone control of male courtship in birds. Horm Behav 54, 227–233. https://doi.org/10.1016/j.yhbeh.2008.04.004.

Gabor, C.R., Bosch, J., Fries, J.N., Davis, D.R., 2013. A non-invasive water-borne hormone assay for amphibians. Amphib Reptil 34, 151–162. https://doi.org/10.1163/15685381-00002877.

Gabor, C.R., Grober, M.S., 2010. A potential role of male and female androgen in species recognition in a unisexual-bisexual mating complex. Horm Behav 57, 427–433. https://doi.org/10.1016/j.yhbeh.2010.01.012.

Gottsberger, B., Gruber, E., 2004. Temporal partitioning of reproductive activity in a neotropical anuran community. J Trop Ecology 20, 271–280. https://doi.org/10.1017/S0266467403001172.

Goymann, W., Landys, M.M., Wingfield, J.C., 2007. Distinguishing seasonal androgen responses from male–male androgen responsiveness—Revisiting the Challenge Hypothesis. Horm Behav 51, 463–476. https://doi.org/10.1016/j.yhbeh.2007.01.007.

Goymann, W., Moore, I.T., Oliveira, R.F., 2019. Challenge Hypothesis 2.0: A Fresh Look at an Established Idea. BioScience 69, 432–442. https://doi.org/10.1093/biosci/biz041.

Grace, J.K., Anderson, D.J., 2014. Corticosterone stress response shows long-term repeatability and links to personality in free-living Nazca boobies. Gen Comp Endocrin 208, 39–48. https://doi.org/10.1016/j.ygcen.2014.08.020.

Groothuis, T.G.G., Carere, C., 2005. Avian personalities: characterization and epigenesis. Neurosci Biobehav Rev 29, 137–150. https://doi.org/10.1016/j.neubiorev.2004.06.010.

Hadfield, J.D., 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmmR Package. J. Stat. Soft. 33, 1–22. https://doi.org/10.18637/jss.v033.i02.

Hau, M., 2007. Regulation of male traits by testosterone: implications for the evolution of vertebrate life histories. BioEssays 29, 133–144. https://doi.org/10.1002/bies.20524.

Hau, M., Gill, S.A., Goymann, W., 2008. Tropical field endocrinology: ecology and evolution of testosterone concentrations in male birds. Gen Comp Endocrin 157, 241–248. https://doi.org/10.1016/j.ygcen.2008.05.008.

Hau, M., Goymann, W., 2015. Endocrine mechanisms, behavioral phenotypes and plasticity: known relationships and open questions. Front Zool 12, S7. https://doi.org/10.1186/1742-9994-12-S1-S7.

Hegner, R.E., Wingfield, J.C., 1987. Effects of Experimental Manipulation of Testosterone Levels on Parental Investment and Breeding Success in Male House Sparrows. The Auk 104, 462–469. https://doi.org/10.2307/4087545.

Herman, R.A., Wallen, K., 2007. Cognitive performance in rhesus monkeys varies by sex and prenatal androgen exposure. Horm Behav 51, 496–507. https://doi.org/10.1016/j.yhbeh.2007.01.005.

Hirschenhauser, K., Winkler, H., Oliveira, R.F., 2003. Comparative analysis of male androgen responsiveness to social environment in birds: the effects of mating system and paternal incubation. Horm Behav 43, 508–519. https://doi.org/10.1016/s0018-506x(03)00027-8.

Hodgson, Z.G., Meddle, S.L., Christians, J.K., Sperry, T.S., Healy, S.D., 2008. Influence of sex steroid hormones on spatial memory in a songbird. J Comp Physiol A 194, 963–969. https://doi.org/10.1007/s00359-008-0369-4.

Hödl, W., 1983. *Phyllobates femoralis* (Dendrobatidae): Rufverhalten und akustische Orientierung der Männchen (Freilandaufnahmen). Begleitveröffentlichungen zu wissenschaftlichen Filmen 30:12-19. Bundesstaatliche Hauptstelle für Wissenschaftliche Kinematographie, Vienna, Austria.

Hunt, K., Wingfield, J.C., Astheimer, L.B., Buttemer, W.A., Hahn, T.P., 1995. Temporal Patterns of Territorial Behavior and Circulating Testosterone in the Lapland Longspur and Other Arctic Passerines. Am Zool 35, 274–284. https://doi.org/10.1093/icb/35.3.274.

Hunt, K.E., Hahn, T.P., Buck, C.L., Wingfield, J.C., 2019. Effect of testosterone blockers on male aggression, song and parental care in an arctic passerine, the Lapland longspur (*Calcarius lapponicus*). Horm Behav 110, 10–18. https://doi.org/10.1016/j.yhbeh.2019.02.001.

Husak, J.F., Fuxjager, M.J., Johnson, M.A., Vitousek, M.N., Donald, J.W., Francis, C.D., Goymann, W., Hau, M., Kircher, B.K., Knapp, R., Martin, L.B., Miller, E.T., Schoenle, L.A., Williams, T.D., 2021. Life history and environment predict variation in testosterone across vertebrates. Evolution 75, 1003–1010. https://doi.org/10.1111/evo.14216.

Kraus, S., Krüger, O., Guenther, A., 2020. Zebra finches bi-directionally selected for personality differ in repeatability of corticosterone and testosterone. Horm Behav 122, 104747. https://doi.org/10.1016/j.yhbeh.2020.104747.

Leary, C.J., 2014. Close-range vocal signals elicit a stress response in male green treefrogs: resolution of an androgen-based conflict. Anim Behav 96, 39–48. https://doi.org/10.1016/j.anbehav.2014.07.018.

Licht, P., Wood, J.F., Wood, F.E., 1985. Annual and diurnal cycles in plasma testosterone and thyroxine in the male green sea turtle *Chelonia mydas*. Gen Comp Endocrin 57, 335–344. https://doi.org/10.1016/0016-6480(85)90212-6.

Lorenzi, V., Earley, R.L., Rodgers, E.W., Pepper, D.R., Grober, M.S., 2008. Diurnal patterns and sex differences in cortisol, 11-ketotestosterone, testosterone, and 17beta-estradiol in the bluebanded goby (*Lythrypnus dalli*). Gen Comp Endocrin 155, 438–446. https://doi.org/10.1016/j.ygcen.2007.07.010.

Marler, C.A., Moore, M.C., 1988. Evolutionary costs of aggression revealed by testosterone manipulations in free-living male lizards. Behav Ecol Sociobiol 23, 21–26. https://doi.org/10.1007/BF00303053.

McGlothlin, J.W., Jawor, J.M., Ketterson, E.D., 2007. Natural variation in a testosterone-mediated trade-off between mating effort and parental effort. Am Nat 170, 864–875. https://doi.org/10.1086/522838.

McGlothlin, J.W., Ketterson, E.D., 2008. Hormone-mediated suites as adaptations and evolutionary constraints. Phil Trans R Soc Lond B 363, 1611–1620. https://doi.org/10.1098/rstb.2007.0002.

Montanarin, A., Kaefer, I.L., Pimentel Lima, A., 2011. Courtship and mating behaviour of the brilliant-thighed frog *Allobates femoralis* from Central Amazonia: implications for the study of a species complex. Ethol Ecol Evol 23, 141–150. https://doi.org/10.1080/03949370.2011.554884.

Moore, I.T., Hernandez, J., Goymann, W., 2020. Who rises to the challenge? Testing the Challenge Hypothesis in fish, amphibians, reptiles, and mammals. Horm Behav 123, 104537. https://doi.org/10.1016/j.yhbeh.2019.06.001.

Moore, I. T., Jessop, T. S., 2003. Stress, reproduction, and adrenocortical modulation in amphibians and reptiles. Horm Behav 43, 39–47. https://doi.org/10.1016/s0018-506x(02)00038-7

Muff, S., Nilsen, E.B., O'Hara, R.B. & Nater, C.R., 2022. Rewriting results sections in the language of evidence. Trends Ecol Evol 37, 203-210. https://doi.org/10.1016/j.tree.2021.10.009.

Mutzel, A., Kempenaers, B., Laucht, S., Dingemanse, N.J., Dale, J., 2011. Circulating testosterone levels do not affect exploration in house sparrows: observational and experimental tests. Anim Behav 81, 731–739. https://doi.org/10.1016/j.anbehav.2011.01.001.

Narins, P.M., Hödl, W., Grabul, D.S., 2003. Bimodal signal requisite for agonistic behavior in a dart-poison frog, *Epipedobates femoralis*. Proc Natl Acad Sci USA 100, 577–580. https://doi.org/10.1073/pnas.0237165100.

Nelson, R.J., 2005. An Introduction to Behavioral Endocrinology. Oxford University Press, New York.

Pašukonis, A., Serrano-Rojas, S.J., Fischer, M.-T., Loretto, M.-C., Shaykevich, D.A., Rojas, B., Ringler, M., Roland, A.B., Marcillo-Lara, A., Ringler, E., Rodríguez, C., Coloma, L.A., O'Connell, L.A., 2022. Contrasting parental roles shape sex differences in poison frog space use but not navigational performance. Elife 11, e80483. https://doi.org/10.7554/eLife.80483.

Peignier, M., Araya-Ajoy, Y.G., Bégué, L., Chaloupka, S., Dellefont, K., Leeb, C., Walsh, P., Ringler, M., Ringler, E., 2022. Exploring links between personality traits and their social and non-social environments in wild poison frogs. Behav Ecol Sociobiol 76, 93. https://doi.org/10.1007/s00265-022-03202-9.

Rasband, W.S., 1997–2021. ImageJ (version 1.50i). U. S. National Institutes of Health. http://imagej.nih.gov/ij/.

Raynaud, J., Schradin, C., 2014. Experimental increase of testosterone increases boldness and decreases anxiety in male African striped mouse helpers. Physiol Behav 129, 57–63. https://doi.org/10.1016/j.physbeh.2014.02.005.

Réale, D., Garant, D., Humphries, M.M., Bergeron, P., Careau, V., Montiglio, P.-O., 2010. Personality and the emergence of the pace-of-life syndrome concept at the population level. Phil Trans R Soc Lond B 365, 4051–4063. https://doi.org/10.1098/rstb.2010.0208.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T., Dingemanse, N.J., 2007. Integrating animal temperament within ecology and evolution. Biol Rev 82, 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x.

Ringler, E., Mangione, R., Ringler, M., 2015. Where have all the tadpoles gone? Individual genetic tracking of amphibian larvae until adulthood. Mol Ecol Resour 15, 737–746. https://doi.org/10.1111/1755-0998.12345.

Ringler, E., Pašukonis, A., Hödl, W., Ringler, M., 2013. Tadpole transport logistics in a Neotropical poison frog: indications for strategic planning and adaptive plasticity in anuran parental care. Front Zool 10, 67. https://doi.org/10.1186/1742-9994-10-67.

Ringler, E., Ringler, M., Jehle, R., Hödl, W., 2012. The female perspective of mating in *A. femoralis*, a territorial frog with paternal care – a spatial and genetic analysis. PLoS one 7, e40237. https://doi.org/10.1371/journal.pone.0040237.

Ringler, E., Szipl, G., Harrigan, R.J., Bartl-Binder, P., Mangione, R., Ringler, M., 2018. Hierarchical decision-making balances current and future reproductive success. Mol Ecol 27, 2289–2301. https://doi.org/10.1111/mec.14583.

Ringler, M., Mangione, R., Pašukonis, A., Rainer, G., Gyimesi, K., Felling-Wagner, J., Kronaus, H., Réjou-Méchain, M., Chave, J., Reiter, K., Ringler, E., 2016. High-resolution forest mapping for behavioural studies in the Nature Reserve 'Les Nouragues', French Guiana. J Maps 12, 26–32. https://doi.org/10.1080/17445647.2014.972995.

Ringler, M., Ringler, E., Magaña Mendoza, D., Hödl, W., 2011. Intrusion experiments to measure territory size: development of the method, tests through simulations, and application in the frog *Allobates femoralis*. PLoS one 6, e25844. https://doi.org/10.1371/journal.pone.0025844.

Ringler, M., Ursprung, E., Hödl, W., 2009. Site fidelity and patterns of short- and long-term movement in the brilliant-thighed poison frog *Allobates femoralis* (Aromobatidae). Behav Ecol Sociobiol 63, 1281–1293. https://doi.org/10.1007/s00265-009-0793-7.

Rodríguez, C., Fusani, L., Raboisson, G., Hödl, W., Ringler, E., Canoine, V., 2022. Androgen responsiveness to simulated territorial intrusions in *Allobates femoralis* males: evidence supporting the challenge hypothesis in a territorial frog. Gen Comp Endocrin 326, 114046, 114046. https://doi.org/10.1016/j.ygcen.2022.114046.

Rodríguez López, C., Hödl, W., 2020. Sound radiation pattern of the advertisement call of the highly territorial poison frog *Allobates femoralis*. Behav Process 170, 103996. https://doi.org/10.1016/j.beproc.2019.103996.

Roithmair, M.E., 1992. Territoriality and male mating success in the dart-poison frog, *Epipedobates femoralis* (Dendrobatidae, Anura). Ethology 92, 331–343. https://doi.org/10.1111/j.1439-0310.1992.tb00970.x.

Romero, L.M., Soma, K.K., O'Reilly, K.M., Suydam, R., Wingfield, J.C., 1998. Hormones and territorial behavior during breeding in snow buntings (*Plectrophenax nivalis*): an Arctic-breeding songbird. Horm Behav 33, 40–47. https://doi.org/10.1006/hbeh.1997.1432.

Rosvall, K.A., Bentz, A.B., George, E.M., 2020. How research on female vertebrates contributes to an expanded challenge hypothesis. Horm Behav 123, 104565. https://doi.org/10.1016/j.yhbeh.2019.104565.

Rosvall, K.A., Bergeon Burns, C.M., Barske, J., Goodson, J.L., Schlinger, B.A., Sengelaub, D.R., Ketterson, E.D., 2012. Neural sensitivity to sex steroids predicts individual differences in aggression: implications for behavioural evolution. Proc R Soc Lond B 279, 3547–3555. https://doi.org/10.1098/rspb.2012.0442.

RStudio Team, 2019. RStudio: Integrated Development for R. Boston. http://www.rstudio.com/.

Schlatt, S., Pohl, C.R., Ehmcke, J., Ramaswamy, S., 2008. Age-related changes in diurnal rhythms and levels of gonadotropins, testosterone, and inhibin B in male rhesus monkeys (*Macaca mulatta*). Biol Reprod 79, 93–99. https://doi.org/10.1095/biolreprod.107.066126.

Schoenemann, K.L., Bonier, F., 2018. Repeatability of glucocorticoid hormones in vertebrates: a meta-analysis. PeerJ 6, e4398. https://doi.org/10.7717/peerj.4398.

Sih, A., Bell, A.M., 2008. Insights for behavioral ecology from behavioral syndromes. Adv Stud Behav 38, 227–281. https://doi.org/10.1016/S0065-3454(08)00005-3.

Sih, A., Mathot, K.J., Moirón, M., Montiglio, P.-O., Wolf, M., Dingemanse, N.J., 2015. Animal personality and state–behaviour feedbacks: A review and guide for empiricists. Trends Ecol Evol 30, 50–60. https://doi.org/10.1016/j.tree.2014.11.004.

Sonnleitner, R., Ringler, M., Loretto, M.-C., Ringler, E., 2020. Experience shapes accuracy in territorial decision-making in a poison frog. Biol Lett 16, 20200094. https://doi.org/10.1098/rsbl.2020.0094.

Stamps, J., Groothuis, T.G.G., 2010. The development of animal personality: relevance, concepts and perspectives. Biol Rev 85, 301–325. https://doi.org/10.1111/j.1469-185X.2009.00103.x.

Stoffel, M.A., Nakagawa, S., Schielzeth, H., 2017. rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol Evol 8, 1639–1644. https://doi.org/10.1111/2041-210X.12797.

Stückler, S., Ringler, M., Pašukonis, A., Weinlein, S., Hödl, W., Ringler, E., 2019. Spatio-temporal characteristics of the prolonged courtship in Brilliant-Thighed Poison Frogs, *Allobates femoralis*. Herpetologica 75, 268–279. https://doi.org/10.1655/Herpetologica-D-19-00010.1.

Těšický, M., Krajzingrová, T., Eliáš, J., Velová, H., Svobodová, J., Bauerová, P., Albrecht, T., Vinkler, M., 2022. Inter-annual repeatability and age-dependent changes in plasma testosterone levels in a longitudinally monitored free-living passerine bird. Oecologia 198, 53–66. https://doi.org/10.1007/s00442-021-05077-5.

Tobler, M., Nilsson, J.-K., Nilsson, J.F., 2007. Costly steroids: egg testosterone modulates nestling metabolic rate in the zebra finch. Biol Lett 3, 408–410. https://doi.org/10.1098/rsbl.2007.0127.

Ursprung, E., Ringler, M., Hödl, W., 2009. Phonotactic approach pattern in the neotropical frog *Allobates femoralis*: a spatial and temporal analysis. Behaviour 146, 153–170. https://doi.org/10.1163/156853909X410711.

Ursprung, E., Ringler, M., Jehle, R., Hödl, W., 2011. Strong male/male competition allows for nonchoosy females: high levels of polygynandry in a territorial frog with paternal care. Mol Ecol 20, 1759–1771. https://doi.org/10.1111/j.1365-294X.2011.05056.x.

van Erp, A.M., Miczek, K.A., 2000. Aggressive behavior, increased accumbal dopamine, and decreased cortical serotonin in rats. J Neurosci 20, 9320–9325. https://doi.org/10.1523/JNEUROSCI.20-24-09320.2000.

Vitousek, M.N., Zonana, D.M., Safran, R.J., 2014. An integrative view of the signaling phenotype: Dynamic links between signals, physiology, behavior and social context. Curr Zool 60, 739–754. https://doi.org/10.1093/czoolo/60.6.739.

Wingfield, J.C., Hegner, R.E., Dufty, A.M., Ball, G.F., 1990. The "challenge hypothesis": theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. Am Nat 136, 829–846. https://doi.org/10.1086/285134.

Wingfield, J.C., Lynn, S., Soma, K.K., 2001. Avoiding the 'Costs' of Testosterone: Ecological Bases of Hormone-Behavior Interactions. Brain Behav Evol 57, 239–251. https://doi.org/10.1159/000047243.

Wingfield, J.C., Ramenofsky, M., Hegner, R.E., Ball, G.F., 2020. Whither the challenge hypothesis? Horm Behav 123, 104588. https://doi.org/10.1016/j.yhbeh.2019.104588.

Table 1. Results of the linear mixed effect model looking at how personality and abiotic factors influence $_{wb}T$ release rates (N = 244 for 39 individuals). Sample size (N) is presented. Results indicating at least weak evidence (*sensu* Muff et al., 2022) are written in bold.

Fixed effects	**Estimate β**	**Standard-Error**	**p-value**
Intercept	5.68	1.45	**< 0.001**
Time of the day (am/pm)	0.06	0.10	0.540
Trial order	-0.00	0.01	0.999
Condition	-0.30	0.08	**< 0.001**
Age (0/1)	-0.11	0.13	0.383
Body size	0.04	0.50	0.934
Calling (yes/no)	0.10	0.11	0.358
Random effects	**Variance**	**Standard-Deviation**	
ID	0.11	0.33	
Residual	0.20	0.45	

Table 2. Covariance and 95% credible intervals between proxies for aggressiveness (agg), exploration (exp) and boldness (bol), and variation in $_{wb}T$ level. Estimates were calculated based on a MCMCglmm model investigating the correlations between the four behaviors measured. An extended version of the table is presented in the Supplementary Table S1.

	among-individual covariance	within-individual covariance
baseline $_{wb}T$ – agg	-0.01 (-0.21, 0.19)	0.08 (-0.09, 0.26)
baseline $_{wb}T$ – bol	-0.05 (-0.28, 0.17)	0.09 (-0.08, 0.27)
baseline $_{wb}T$ – exp	0.18 (-0.01, 0.45)	-0.14 (-0.31, 0.03)
Δ $_{wb}T$ – agg	-0.02 (-0.14; 0.09)	0.01 (-0.19; 0.22)
Δ $_{wb}T$ – bol	0.02 (-0.14; 0.16)	-0.18 (-0.39; 0.03)
Δ $_{wb}T$ – exp	-0.05 (-0.21; 0.1)	0.11 (-0.1; 0.32)

Figure 1. Experimental design. We repeatedly sampled $_{wb}T$ under two following conditions: without the use of any acoustic stimuli ('baseline') and immediately after a consecutively deployed suite of behavioral tests ('experimental'). We aimed for obtaining three replicates of both 'baseline' and 'experimental' hormonal samples per individual (a total of six measurements per frog). Half of the tested individuals started with the 'baseline' condition (as visualized in the figure), while the other half started with the 'experimental' condition. Consecutive samples were always taken more than 24 hours apart. After every second trial we added a break of at least 3 days to minimize any confounding effects of the procedure on the measurements.

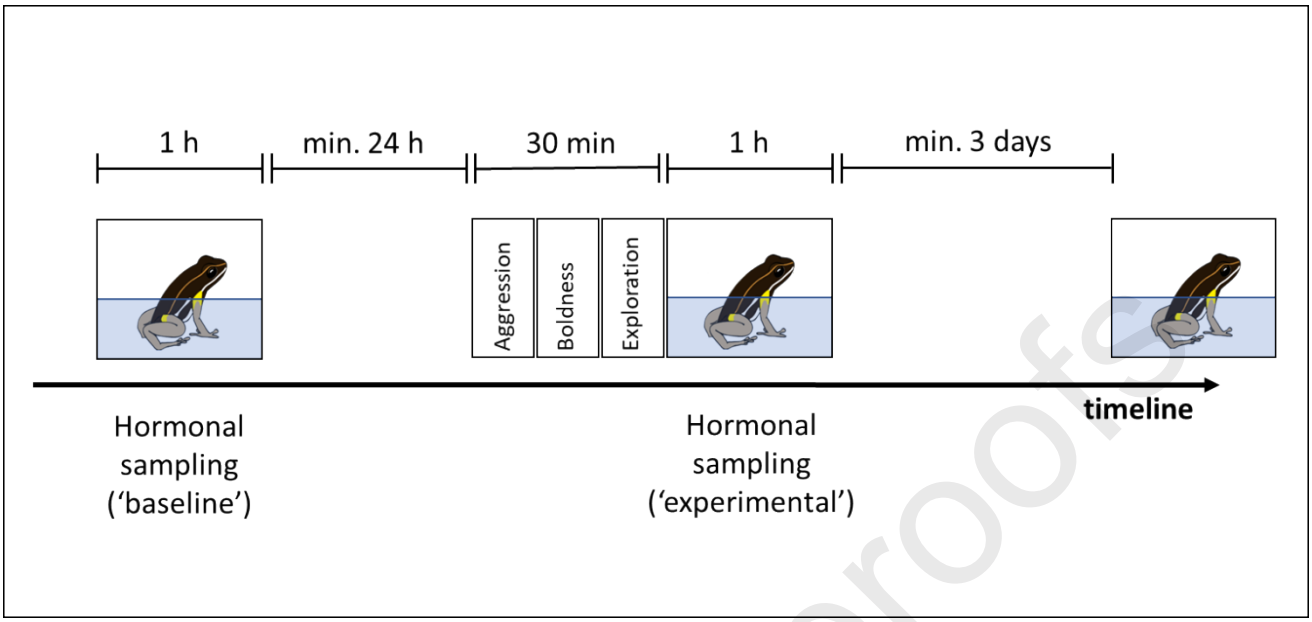
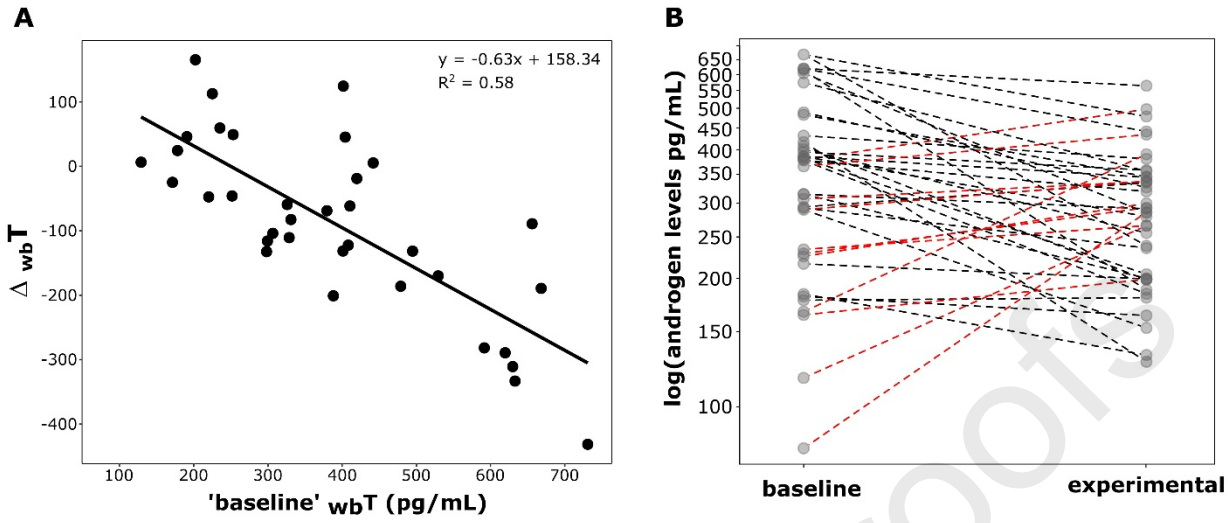


Figure 2. Comparison of average individual $_{wb}T$ levels from baseline samples and samples taken immediately after behavioral experiments. A) Correlation between average baseline $_{wb}T$ and the relative change $\Delta_{wb}T$ after behavioral testing; B) Pairwise comparisons of individual males. Red lines indicate individuals for which average $_{wb}T$ measured in the 'experimental' condition is increased compared to 'baseline'. Black lines indicate individuals for which the average $_{wb}T$ level is decreased in the 'experimental' condition compared to 'baseline'.



Highlights

- Hormones play a key role in modulating social behavior in animals
- Baseline water-borne testosterone levels were highly repeatable among males
- Testosterone levels predicted individual exploratory tendency