General and Comparative Endocrinology 300 (2021) 113638

Contents lists available at ScienceDirect

General and Comparative Endocrinology

journal homepage: www.elsevier.com/locate/ygcen

Research paper

Excreted testosterone and male sexual proceptivity: A hormone validation and proof-of-concept experiment in túngara frogs

Alexander T. Baugh *, Sophie L. Gray-Gaillard

Department of Biology, Swarthmore College, 500 College Avenue, Swarthmore, PA 19081, United States

ARTICLE INFO

Keywords:
Creatinine
GnRH
hCG
Non-invasive hormones
Phonotaxis
Testosterone

ABSTRACT

Conventional methods for sampling hormones often preclude strong inference experimental designs, including repeated measures of both hormones and behavior and balanced or simultaneous designs for hormone-behavior sampling. In amphibians there is an opportunity to non-invasively and repeatedly sample excreted steroids in the water. We examined testosterone (T) in túngara frogs (*Physalaemus (=Engystomops) pustulosus*) using minimally invasive water-borne methods. First, we validated procedures for the collection, extraction and measurement of T in adult males and females using pharmacological challenges coupled with estimates of parallelism and recovery determination. Next, we extended the timeline of sampling over 9 days in order to evaluate the kinetics of excretion (baseline phase, challenge phase, recovery phase), including the estimation of individual differences during baseline sampling. We also estimated concentrations of creatinine (Cr) in each water sample and evaluated whether correcting for this proxy of urine concentration significantly decreased error variance in T estimates. Lastly, we incorporated a standardized and repeated measures assay of male sexual proceptivity (phonotaxis) during the predicted peak T and recovery T timepoints. We found strong evidence supporting the utility of these methods for precise, biologically informative estimates of T in both sexes. Males had higher T than females and responded to pharmacological challenges by elevating T substantially within 48 h of challenge (hCG, GnRH). Males exhibited repeatability in baseline T and phonotaxis frequencies were positively associated with higher T. Adjusting T levels for the simultaneous measure of Cr significantly improved model fit, which in conjunction with marked variation in urine concentration, suggests that urine likely serves as the major source of excreted T. In summary, this proof-of-concept and methods study demonstrates the utility and accuracy of measuring water-borne T and behavior in amphibians.

1. Introduction

Conventional collection of tissue samples for the measurement of circulating hormones can directly impact subsequent secretion and behavior (Romero and Reed, 2005; Baugh et al., 2013, 2017a) and provides only an instantaneous snapshot of endocrine status, which may not reflect longer-term, more integrative measures (Good et al., 2003; Touma and Palme, 2005; Bastien et al., 2018; Baugh et al., 2018). Circumventing these problems often demands unbalanced designs that position hormone sampling as the final intervention. Such designs can limit our understanding of hormone–behavior relationships which can be bidirectional and rapid (Gabor and Grober, 2010), including for products in the hypothalamic-pituitaryadrenal (HPA) (Breuner and Wingfield, 2000; Baugh et al., 2017a; reviewed in Leary and Baugh, 2020) and the hypothalamic-pituitary-gonadal (HPG) axes (Cheng, 1992, Trainor et al., 2007, Trainor et al., 2008). Thus, sampling methods that minimize or eliminate the influence of the experimenter on the animal's experience have the potential to improve our ability to draw inferences about causation by allowing the simultaneous measurement of both trait categories (Still et al., 2019). Additionally, these approaches also permit repeated or continuous monitoring of both endocrine status and behavior, thus allowing simultaneously for a more granular as well as temporally integrative phenotyping of hormone action, estimates of individual variation and reaction norm approaches. Improved estimates of individual (co)variation in hormone concentrations could thus help resolve some inherent challenges of understanding complex endocrine systems that exhibit hyper-labile traits (Dingemanse et al., 2010).

A few studies have recently validated urinary steroid methods for amphibians, especially glucocorticoids (reviewed in: Kindermann et al.,

* Corresponding author at: Swarthmore College, 500 College Avenue, Singer Hall, Swarthmore, PA 19081, United States.
E-mail address: abaugh1@swarthmore.edu (A.T. Baugh).

https://doi.org/10.1016/j.ygcen.2020.113638
Received 22 July 2020; Received in revised form 9 September 2020; Accepted 28 September 2020
Available online 2 October 2020

2012; Narayan, 2013). Urine sampling has the advantage of being less invasive than tissue collection and yet can provide adequate concentrations for estimating multiple steroids from the same sample. Urinary sampling, however, is often limited to adults of larger species and typically involves animal handling (via cloacal sampling). Handling stress may therefore interfere with subsequent behavior or hormone secretion or both. In contrast, water-borne hormone sampling methods have the potential to be used with minimal or no animal handling and for any size of amphibian, including tadpoles (reviewed in Narayan et al., 2019). One of the potential limitations of water-borne methods is the fact that there are potentially multiple sources of excretion, which may vary in importance and metabolite makeup.

In contrast to the increasing number of studies of water-borne CORT (reviewed in Narayan et al., 2019) and estradiol and progesterone (Baugh et al., 2018), there are currently no validated methods for the measurement of water-borne androgens. Here we validated water-borne testosterone (T) in túngara frogs (*Physalaemus (=Engystomops) pustulosus*) and use these procedures to examine an understudied aspect of reproductive endocrinology—the androgen correlates of proceptive behavior in males. Our validation and proof-of-concept study consisted of several components, including: (1) methodologies for the collection, extraction and measurement of T in adult males and females using pharmacological challenges with human chorionic gonadotropin (hCG) and gonadotropin releasing hormone (GnRH) coupled with estimates of parallelism and recovery determination; (2) a 9-day repeated sampling timeline to evaluate the kinetics of excretion (baseline, challenge, and recovery phases); (3) estimating individual differences during baseline; (4) we co-sampled creatinine (Cr) in each water sample and evaluated whether statistically correcting for Cr improved the T estimates because urine may serve as a major source of excreted water-borne T; and (5) we conducted a proof-of-concept experiment using a repeated measures evaluation of a standardized assay of male sexual proceptivity (phonotaxis). We predicted the peak and return-to-baseline T timepoints during the 9-day experimental timeline and tested male phonotaxis during these phases, thereby allowing us to also test the novel hypothesis that the frequency of this male sexual behavior is positively related to the integrated measure of excreted T.

Túngara frogs have been studied widely in the context of the ecology and evolution of sexual communication (Ryan et al., 2019) and increasingly to understand the endocrine basis for sexual behavior. This work has been conducted using wild-caught adults (Baugh and Ryana, 2010, 2017, Chakraborty and Burmeister, 2009, 2010, Lynch and Wilczynski, 2005, 2006; Marler and Ryan, 1996) and lab-reared populations (Baugh et al., 2012, 2018; Baugh and Ryan, 2010b; Lynch et al., 2006). Túngara frogs excrete hormones including their metabolites into the surrounding water (Baugh et al., 2018), as has been demonstrated in a few other amphibians (reviewed in Narayan et al., 2019). During the breeding season adult males and females typically spend several hours at night immersed in water while advertising vocally and hydrating/ovipositing their clutches, respectively (Ryan, 1985). We therefore assume that immersion in water is likely an important environmental stimulus during sexual readiness and thus naturally congruent with hormone sampling. We used a 60-min duration water bath, which has been shown to reliably capture physiologically relevant concentrations of other excreted steroid hormones, including corticosterone, 17-β estradiol, and progesterone in this species (Baugh et al., 2018).

Here we extend that previous work focusing on T. The current study consisted of two phases in which frogs were placed in water baths to quantify water-borne T concentrations after pharmacological challenge (or control). We validated methods for collection, extraction and quantification of physiologically relevant T and examined the utility of correcting for an index of excreted urine concentration (via levels of creatinine, hereafter Cr). We further estimated sex differences and the repeatability of excreted T, as well as evaluated the link between male sexual behavior (phonotaxis) and excreted T.

2. Methods

2.1. The system

We used lab-reared adult túngara frogs that were the first generation of offspring derived from breeding adults collected in July 2017 in the field from a population near Gamboa, Panama. All animals were maintained at 27C, ambient humidity and a 12:12 L:D cycle in terraria in small single sex groups (N = 5–10) until assigned to treatments. Once assigned, each frog was measured for mass using a digital balance to the nearest 0.01 g (mean ± SD: 0.74 ± 0.23 g) and, in Phase 2 also body length (Snout-to-Vent Length, SVL) using calipers to the nearest 0.01 mm (mean ± SD: 22.60 ± 1.90 mm). Frogs were then housed alone in labeled small plastic containers (6 × 5 cm, diameter × height) with moist sphagnum moss and ventilation and supplied daily with *ad libitum* fruit flies dusted with a vitamin powder (Repashy Calcium-plus). In lieu of individual identifiers (e.g. via toe clipping), we maintained frogs in labeled containers and water baths.

2.2. Experimental design

2.2.1. Phase 1

In February of 2018 male (N = 16) and female (N = 26) frogs were assigned randomly to either experimental (human chorionic gonadotropin, hCG; 500 IU g^{-1}; hCG; Sigma Cat. No. C1063) or control (vehicle, 0.9% physiological saline) treatments in equal numbers per sex. This dose of hCG has been previously validated for effectively inducing gonadal steroid elevations (Baugh et al., 2018) and species-typical phonotaxis (Baugh and Ryan, 2010b; Lynch et al., 2006). Frogs were maintained individually during the experiment and for 24 h prior to the first sample collection. Each individual was sampled for water-borne T for 60 min prior to injection (pre-injection) and again 24 h after injection (post-injection). Injections were performed intraperitoneally using a 0.3 cc insulin syringe (BD Ultra-Fine) fitted with a 31-gauge needle (Baugh and Ryan, 2010b; 2017).

2.2.2. Phase 2

In July of 2019 we used males to extend this study. We used similar pharmacological challenges, including the same route of administration, the same dose of hCG (500 IU g^{-1}; N = 10) and vehicle (0.9% physiological saline; N = 10) and we incorporated an injection of gonadotropin releasing hormone (GnRH, 0.5 μg g^{-1}; N = 10; Sigma Cat. No. L8008) as a treatment. Whereas hCG acts as an a luteinizing hormone (LH)-like hormone binding to LH-receptors in the gonads, GnRH acts one stage upstream in the HPG axis as the luteinizing hormone releasing hormone that binds to receptors in the pituitary. Though hCG has been used extensively in this species, this is the first use of GnRH. This dose has been shown to maximally induce sperm release in male Booroolong frogs (*Litoria booroolongensis*; Silla et al., 2019) and is similar to a validated dose (0.4 μg g^{-1}) used to induce amplexus behavior in Northern leopard frogs (*Lithobates pipiens*; Vu et al., 2017). *Phase 2* extended the timeline of sampling in order to better understand the temporal dynamics of excretion patterns, including *baseline*, *acute* and *integrated* responses to challenge and the *recovery* to baseline (Fig. 1). Moreover, we split each water sample for the independent measurement of T and Cr. Urinary steroid concentrations are often standardized to levels of Cr—a by-product of muscle metabolism—to correct for water content. Because we predicted that an important fraction of water-borne steroids is contributed via urinary excretion, and because whether and how much a frog urinates can be variable, we wanted to correct our T estimates for variation in urinary Cr (Narayan et al., 2010) and compare this to uncorrected T levels to evaluate the value of this additional procedure. Additionally, we sampled baseline T daily for three days prior to challenge to estimate the repeatability of T excretion as an indicator of the reliability of this protocol and because hormonal repeatability is itself of interest biologically (Hau et al., 2016). Further, we performed a

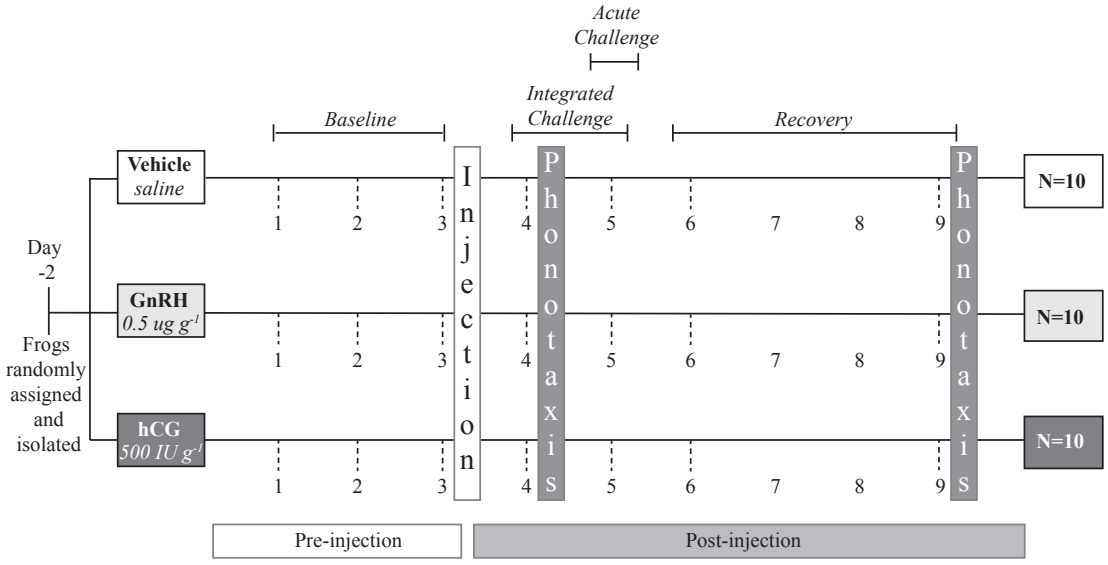


Fig. 1. Experimental timeline for Phase 2. Water baths collected on Days 1–6, and 9 at the vertical dashed lines.

validated test of sexual behavior (phonotaxis) during what we predicted would be peak T and again after what we predicted would be a return to baseline T in order to demonstrate that behavior-hormone sampling designs can be successfully executed and to test for the novel association between T and male phonotaxis in anurans. Lastly, we also validated other canonical aspects of the physiological procedures, including parallelism, recovery, coefficients of variation and drying methods.

2.3. Water sample collection

Individual frogs were placed into translucent and disposable Dixie™ cups (polystyrene, 295 mL capacity) holding 25 mL of 'frog water' and a second cup placed above, allowing the frogs headspace but ensuring the frogs were immersed in the water during the entire 60-min period (Supplementary figure 1). Following Odum and Zippel (2008), we prepared 'Frog water' with the following dissolved solutes per 30 L of RO water: 1.2 g CaCl2, 1.38 g MgSO4, 1.08 g KHCO3, 0.9 g NaHCO3 and 0.038 g of a commercial trace-element mixture (#6 chelate trace element from Homegrown Hydroponics). Following the 60-min sample collection, frogs were returned to their individual containers and water samples were course filtered (VWR Grade 417) to remove large particles and transferred to 50 mL conical centrifuge tubes and then maintained at −80C for up to 7 months prior to extraction. Control water samples (i.e., no frog) were collected to evaluate possible contamination (all control samples registered below the detection limits of the assay). Sampling was conducted between 0800 and 1200 to control for diel rhythms in hormone secretion.

2.4. Sample processing, parallelism and recovery

2.4.1. Extraction

Water samples were thawed at 4C overnight. In *Phase 1* we extracted the full 25 mL of individual water samples to quantify T, whereas in *Phase 2* we first vortexed the water sample and split it into two equal aliquots for the separate measurement of T and Cr. Water samples were extracted using solid phase columns (SPE; Sep-Pak C18, 50 mg sorbent; Waters Corp., Milford, MA) placed in a 24-port vacuum manifold (United Chemical Technologies, LLC, Bristol PA). SPE has been shown to result in reliable analyte recovery and remove interfering compounds (Lee and Goeger, 1998; Newman et al., 2008) and has been used successfully for amphibian water-borne hormones (Gabor et al., 2013a, 2013b, 2015; Baugh et al., 2018). Columns were first activated with 4 mL of methanol (ACS grade) and equilibrated with 4 mL RO water. Samples were then drawn through the columns under 15 bar of vacuum pressure at a rate of approximately 10 mL per minute. An additional 4 mL of RO water was then processed through each column. We slowly eluted each column with 3 mL of methanol (HPLC grade) into borosilicate vials. Water subsamples for Cr were not extracted, but instead dried directly. We did not sample blood for the measurement of plasma T concentrations because doing so in this species typically involves the use of an invasive cardiac puncture technique (Baugh et al., 2018).

2.4.2. Drying

Elutants were dried using a centrifuge concentrator (Savant SpeedVac SPD1010, Thermo Fisher Scientific, Waltham, MA) at 45C for 2–4 h and frozen at −20C until reconstitution (Baugh et al. 2018). Approximately 12 h prior to assaying, samples were reconstituted in a mixture of 95% commercial assay buffer and 5% reagent grade ethanol. In *Phase 1*, in which all 25 mL of water sample was used for testosterone analysis, female samples were reconstituted to 150 µL and all male samples were reconstituted to 250 µL. These sex specific dilution factors were calibrated empirically by collecting additional samples from non-experimental animals, pooling them in male and female pools, reconstituting to various volumes and conducting the enzyme immunoassay (EIA) for T to identify dilutions that reliably fell in the linear portion of the standard curve. In *Phase 2*, in which water samples (from males only) were halved for the measurement of T and Cr, we maintained this dilution factor (1:1) by reconstituting the halved water sample in 125 µL of the 95%/5% buffer/alcohol mixture.

We used the SpeedVac to dry the paired (halved) water samples for Cr. Prior to drying the experimental aliquots, we compared two methods for drying water samples for Cr estimation using paired split samples (25 mL water bath per frog, split into 12.5 mL aliquots). We used five split samples in total, four came from 60-min water baths for each of four unique adult males (not used in other parts of this study) and one control sample that was treated identically but had not housed a frog. Method 1 used a single borosilicate vial per unique sample (12.5 mL total volume), first filled with a 3-mL aliquot and dried and then refilled and dried successively three more times for a total of four rounds of drying per vial/sample. The vials were then reconstituted as described previously. Method 2 used four borosilicate vials per unique sample (3 mL water each) which were dried simultaneously and then reconstituted individually and pooled into a single vial at the end. Because these two methods proved indistinguishable (see Results) we opted to use Method 1 for all experimental samples.

2.4.3. Parallelism & recovery

2.4.3.1. Testosterone. In *Phase 2* a 0.5 mL subsample of each experimental water sample (25 mL) was pooled and split into aliquots to perform parallelism and recovery analysis. For parallelism analysis, a 25 mL portion of the pooled sample was serially diluted at 1:1 (i.e. 250 μL with 95% assay buffer and 5% reagent alcohol), 1:2, 1:4, and 1:8 and assayed. For recovery, two 12 mL aliquots of the pooled samples were stripped of endogenous steroids using 7 mg mL^{-1} of dextran-coated activated charcoal (Sigma Cat. No. C6241). Samples were vortexed, incubated at 37C for 4 h, centrifuged for 20 min (4000 RCF) under refrigeration (4C), and supernatants were collected (Baugh et al., 2018; Delehanty et al., 2015). Centrifugation and supernatant collection were repeated for a total of three rounds. One stripped/unspiked sample went through the extraction process and was plated at 1:1. The second stripped sample was spiked with 640 pg mL^{-1} of purified T (supplied by EIA kit), extracted, and plated. A 12 mL sample of the pool remained unstripped/unspiked and was extracted and plated at 1:1. We conducted one additional analysis of recovery by preparing kit-supplied standards 3 (1,600 pg mL^{-1}), 4 (640 pg mL^{-1}), and 5 (256 pg mL^{-1}) in blank frog water, extracting the samples, and reconstituting the samples in 300 μL assay buffer (i.e. this method did not involve any stripping of actual samples).

2.4.3.2. Creatinine. For parallelism analysis, a 50 mL pooled water sample was dried and reconstituted using a serial dilution at 2:1 (130 μL water), 1:1 (260 μL water), 1:2, and 1:4 and assayed as described below. Additionally, we prepared five replicate 2.5 mL aliquots of a separate pooled sample, dried them and reconstituted them at a dilution factor 2:1.

2.4.4. Enzyme immunoassay for testosterone

We estimated steroid concentrations using commercial EIA kits (DetectX® kits, Arbor Assays, Ann Arbor, MI) for water-borne testosterone (Cat. No. K032, Goat anti-Rabbit IgG). Reconstituted samples and kit reagents were allowed to reach room temperature prior to use and samples were vortexed prior to plating. Samples were randomly assigned to wells and assayed in duplicate along with blanks, standards, stripped samples, and stripped/spiked samples. Samples were assayed following manufacturer instructions. Briefly, 50 μL of sample or standard were plated into wells along with conjugate and antibody. Plates were then placed on an orbital shaker (500 RPM) at room temperature for 2 h and then washed four times with wash buffer (supplied by kit). Substrate was then added and the plate was incubated at room temperature for 30 min without shaking. The reaction was stopped and optical densities were read at 450 nm on a $Versa_{max}$ microplate reader with SoftMax Pro software using a four-parameter curve fitting equation (Molecular Devices, Sunnyvale CA). Intra- and inter-assay coefficients of variation (CV) for the assays were estimated by including a minimum of two stripped/spiked samples per plate, thereby incorporating cumulative technical error during extraction and assaying (see Baugh et al., 2018). We accepted the average of duplicate wells. The assay has a sensitivity and detection limit, respectively, of 9.92 pg mL^{-1} and 30.6 pg mL^{-1}. The cross-reactivity of the antiserum is as follows: 100% for T, 56.8% for 5a-dihydrotestosterone (hereafter DHT), and 0.27% for androstenedione. The high cross reactivity of the T antibody for DHT could indicate that our T estimates may better represent testosterone plus DHT (i.e. 'androgens'). However, a recent study using high performance liquid chromatography-mass spectrometry showed that water-borne DHT concentrations are undetectable in túngara frogs including adults that were injected with a 500 IU g^{-1} dose of hCG and sampled 24 h later (Baugh et al., 2018). Therefore, we hereafter provisionally refer to these estimates as testosterone.

2.4.5. Assay for creatinine

Creatinine was assayed directly without extraction using the Jaffe reaction method, following the manufacturer's instructions (DetectX® kit, Arbor Assays, Ann Arbor, MI; Cat. No. K002). The Cr parallelism component indicated that a reconstitution volume of 65 μL was suitable for the average water sample and therefore dried extracts of each sample were reconstituted at 65 μL with RO water. We randomly assigned samples to wells. One sample was lost due to assay plating error. Each plate was run with duplicates of a blank and standards for the standard curve, and samples were run in singlet, which was necessary given the parallelism results to achieve an adequate Cr concentration given the splitting of frog water samples for the simultaneous measure of T and Cr. Replicates of two standards from the standard curve (0.05 mg mL^{-1} and 0.025 mg mL^{-1}) were plated in triplicate to calculate intra- and inter-assay CVs. Plates were read at 490 nm on a $Versa_{max}$ microplate reader using the kinetic read function during a 3.5 h period with SoftMax Pro software using a four-parameter curve fitting equation (Molecular Devices, Sunnyvale CA). Readings were taken every 10 min following 300 s of shaking. Prior to assaying of experimental samples, we validated these microplate reader parameters using non-experimental samples distributed across two plates. We found that optical density readings increased over the course of 3 h and leveled off at approximately 3.5 h. Therefore, we accepted the average values during the 3.5 h kinetic read. The detection limit and sensitivity of the assay are 0.00037 mg mL^{-1} and 0.00019 mg mL^{-1}, respectively.

2.5. Acoustic stimuli and behavioral testing

During the breeding season, male túngara frogs vocally compete and attract female mates with a species-specific vocal advertisement. The vocalization is known as the simple call or the 'whine' which can be enhanced with suffixes known as 'chucks' to produce a complex call known as a 'whine-chuck' (Ryan, 1985). Adult females exhibit positive phonotaxis toward these calls and prefer the complex whine-chuck calls over simple calls in laboratory and field settings, using natural and synthetic stimuli (Ryan and Rand, 2003; Bernal et al., 2009; Baugh and Ryan, 2010b). Similarly, male túngara frogs, including wild-caught and lab-reared adults and juveniles exhibit positive phonotaxis toward these synthetic and natural advertisement calls (Baugh et al., 2012; Baugh and Ryan, 2010a,b). Several previous studies have identified endocrine correlates of phonotaxis in females (Baugh et al., 2012, 2018, Baugh and Ryan, 2010a,b; Chakraborty and Burmeister, 2009, 2010; Lynch et al., 2006; Lynch and Wilczynski, 2005, 2006; Marler and Ryan, 1996) but only one study to date has explored the hormonal basis of this proceptive behavior in males—Baugh and Ryan (2017) documented that the neuropeptide arginine vasotocin positively modulates male phonotaxis.

To further examine the link between phonotaxis in males and excreted T, males were tested in standardized two-choice phonotaxis trials during the dark period of their light cycle. Experiments were conducted inside a rectangular, sound-attenuating chamber (2.4 × 1.8 × 1.78 m, L × W × H; Industrial Noise Control, North Aurora, IL) under infrared lighting and outfitted with acoustic tiles on walls and ceiling and carpeted with low pile carpet inside a temperature-controlled (~27 °C) laboratory. Playbacks were performed using a desktop PC (Dell 5520, Windows 7, with a Nvidia Quadro M1200 sound card calibrated by Engineering Design, Berkeley, CA) using SIGNAL 5.0 (Engineering Design), and connected to two single-channel high fidelity gain control potentiometers (SPL Electronics GmbH, Niederkrüchten, Germany), a two-channel power amplifier (Crown XLS 1000; Crown Audio Inc., Elkhart, IN), and two satellite speakers (Mod1, Orb Audio, NY, NY) located 2 m apart at opposite ends of the sound chamber at equal distances from the center of the chamber. The sound pressure level (SPL) of stimuli were calibrated to 82 dB (LCF_{max}; SPL re 20 μPa) at 1 m using a SoundTrack LXT sound pressure level meter (Larson Davis, Provo, UT), which approximates a natural call amplitude in this species (Ryan, 1985), with its calibrated microphone pointed directly toward each

speaker from the central release point at the center of the chamber. Frog movement was recorded and monitored live via a wide-angle infrared camera (Ikegami ICO-49, Japan) mounted from the center of the sound chamber ceiling and connected to a desktop PC (Dell 5520, Windows 7) located outside the chamber operating Ethovision XT (Version 9, Noldus, Wageningen, NL).

We used two synthetic stimuli throughout this study, which were matched for peak amplitude before playback: a simple whine and the whine with one chuck appended to the end. The chuck is twice the peak amplitude of the whine. The whines in these signals consist of only the fundamental frequency; it has been shown previously that the upper harmonics of the whine do not influence mate choice in the laboratory (Ryan and Rand, 1990; Rand et al., 1992), and that these synthetic conspecific signals are as attractive as natural signals (Ryan, unpublished data). We synthesized these stimuli based on the mean values for the parameters of the calls in the population by shaping sine waves using custom software (J. Schwartz, Pace University at Pleasantville, NY, U.S.A.; sampling rate 20 kHz and 8 bit depth). We calculated mean values for the population based on the calls from 50 males recorded in July 1996 with a Marantz PMD 420 recorder and a Sennheiser ME 80 microphone with K3U power module on magnetic cassette tape. Additional information on the call parameters used, the synthesis procedure, and spectrograms and waveforms of these two signals can be found in Ryan et al. (2003).

In order to minimize any potential side bias in the chamber or first caller preference (Bosch and Márquez, 2002), we randomly assigned the order of frogs, the order of stimulus onset (leading versus lagging speaker), and the location of the stimuli. We broadcast the calls antiphonally at the average call rate in the population of origin (1 call/speaker/2 sec; Ryan, 1985). Each male was tested in a single phonotaxis trial on Day 4 and Day 9 immediately following each animal's water bath sample collection. Each frog was placed in the center of the sound chamber under an acoustically transparent funnel, and the trial started following an initial 10 s of playback. If the frog entered a semicircle choice zone that extended 10 cm from the face of either speaker, a 'choice' (C) was scored and we recorded which stimulus was chosen, and then terminated the trial. If the frog did not enter a choice zone after 5 min, we scored the trial as a 'no choice' (NC).

2.6. Statistical analysis

2.6.1. Phase 1

Testosterone concentrations are expressed as picograms per milliliter per hour (pg mL^{-1} hr^{-1}; Baugh et al. 2018) and T levels were log-transformed to improve the normality of error variances.

We used a general linear model with a within-subjects term (timepoint) and between-subjects term (injection treatment) with a full factorial design.

2.6.2. Phase 2

Testosterone concentrations are reported as picograms per milliliter (of buffer) per hour per millimeter of body length (pg mL^{-1} hr^{-1} mm^{-1}_{SVL}; Baugh et al., 2018). Testosterone and Cr concentrations were natural log-transformed to improve the normality of error variances. We constructed and compared three linear mixed effects models using maximum likelihood estimates in order to evaluate whether adjusting T concentrations for Cr levels increased the explanatory power. All three models included repeated measures for all seven days of natural log-transformed T sampling with subject as a random factor and used Type 3 tests for fixed effects. Model 1 was a main effects model with fixed effects for Day (7 levels) and Treatment (3 levels). Model 2 added the interaction term (Day*Treatment) and Model 3 then added log-creatinine as a covariate (i.e. adjusted T concentrations). We compared the log-likelihood and AICc estimates among models to evaluate explanatory power. Lastly, we performed four *a priori* orthogonal pair-wise comparisons to compare predicted values of T (i.e. T corrected for Cr levels) between treatment groups, as follows: (a) *Baseline* (Days 1–3); (b) *Integrated Challenge* (Days 4–5); (c) *Acute Challenge* (Day 5); (d) *Recovery* (Days 6–9). Of the 210 possible observations (i.e. 30 subjects for 7 days) there were four missing observations for T and one for Cr (final k = 205–206) owing to those samples falling below the detection limits. We report uncorrected p-values for these *a priori* comparisons, though note that the statistical significance (at $\alpha = 0.05$) of the results would be unchanged using conventional corrections for multiple comparisons (e.g. Sidak). We attempted to analyze the phonotaxis choice data using generalized linear mixed effects models. However, because the odds ratios were inestimable, likely owing to the small sample size (N = 10/treatment), we instead report the descriptive results.

Finally, we report the repeatability of T concentrations using a mixed effects approach to estimating intraclass correlations (*ICC*), including 95% confidence intervals. We also calculated the within-subject variance (σ_w) along with 95% CI. When 95% CI do not overlap zero this outcome is interpreted as statistically significant (Dingemanse and Dochtermann, 2013; Baugh et al., 2017b). We used Days 1–3 (*Baseline*) for this analysis because it precedes any experimental treatment and provides the largest sample size (N = 30; k = 88 observations owing to two T samples that were below the assay's detection limit). We performed two ICC calculations, one simply using ln-testosterone concentrations (i.e., agreement repeatability) and the other using the predicted values from the ln-testosterone model that specified ln-creatinine as a covariate (i.e. adjusted repeatability). We used a comparison of slopes *t*-test to evaluate parallelism. All statistical analyses were performed in SPSS (Version 21, IBM, Armonk, NY) and met parametric assumptions (normality of error, equality of (co)variances, sphericity) unless noted otherwise.

3. Results

3.1. Phase 1

Concentrations of T in hCG-injected males increased from pre- to post-injection timepoints (Fig. 2), owing to a unanimous increase among males (Supplementary figure 2). In contrast, hCG-treated females and control groups of both sexes exhibited no change in average T concentrations across timepoints (Fig. 2; Supplementary figure 3). Statistically,

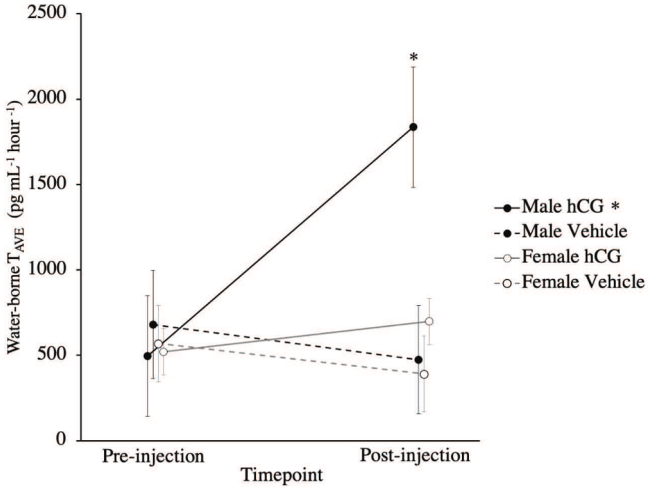


Fig. 2. Phase I: Water-borne T concentrations (mean ± 1 SE) for hCG (500 IU g^{-1}) and vehicle (saline) injected male and female repeated measures treatment groups. Water samples were taken at two timepoints 24 h apart (N_{hCG} = 8 males, 13 females; $N_{Vehicle}$ = 8 males, 13 females). Lines indicate average concentration changes for each group across time point. For illustration purposes the x-axis intercepts are jiggered. * indicates significant ($p < 0.05$) elevation compared to vehicle treatment.

there was a main effect of sex with males having a higher concentration of T than females ($F_{1,34} = 4.28$, $p = 0.046$). Males at the pre-injection timepoint did not differ between hCG and control treatments ($F_{1,34} = 0.31$, $p = 0.583$), while hCG treated males had significantly higher T concentrations than vehicle treated frogs at the post-injection timepoint ($F_{1,34} = 19.04$, $p < 0.001$). In the hCG-treated frogs, males exhibited a greater increase in T following injection than females ($F_{1,34} = 6.19$, $p = 0.018$). This increase after injection in hCG-treated males remained statistically significant even when the male that experienced the greatest increase was removed ($t_5 = 4.91$, $p = 0.004$, two-tailed; Supplementary figure 2). There was no difference between sexes in vehicle injected frogs ($F_{1,34} = 0.19$, $p = 0.665$). Females at the pre- and post-injection timepoint, respectively, showed no significant difference among treatment groups ($F_{1,34} = 0.034$, $p = 0.855$; $F_{1,34} = 1.66$, $p = 0.206$). The serial dilution of pooled samples of the male hormone extracts resulted in parallelism (Supplementary figure 4). The intra-assay CV was 12.7%.

3.2. Phase 2: Pharmacological challenges in males

3.2.1. Creatinine-adjusted testosterone estimates

Baseline levels of T were consistently and uniformly low among the three treatment groups for the three days prior to injection. After injection, T levels spiked, following a lag of 24–48 h, in the GnRH and especially the hCG treatment groups. Both groups recovered to baseline levels by Day 9. In addition to the effect of Day and Treatment, there was considerable variation in T (see *Repeatability*) and Cr concentrations among subjects (Fig. 3; also see Supplementary figure 5). And Cr concentrations were positively correlated with T concentrations for each day of sampling (Pearson's *r* range: 0.38 – 0.73; Supplementary figure 6). Lastly, Cr concentrations increased incrementally across the sampling period (Supplementary figure 7). These three observations supported the potential utility of adjusting T estimates using each sample's co-sampled Cr level.

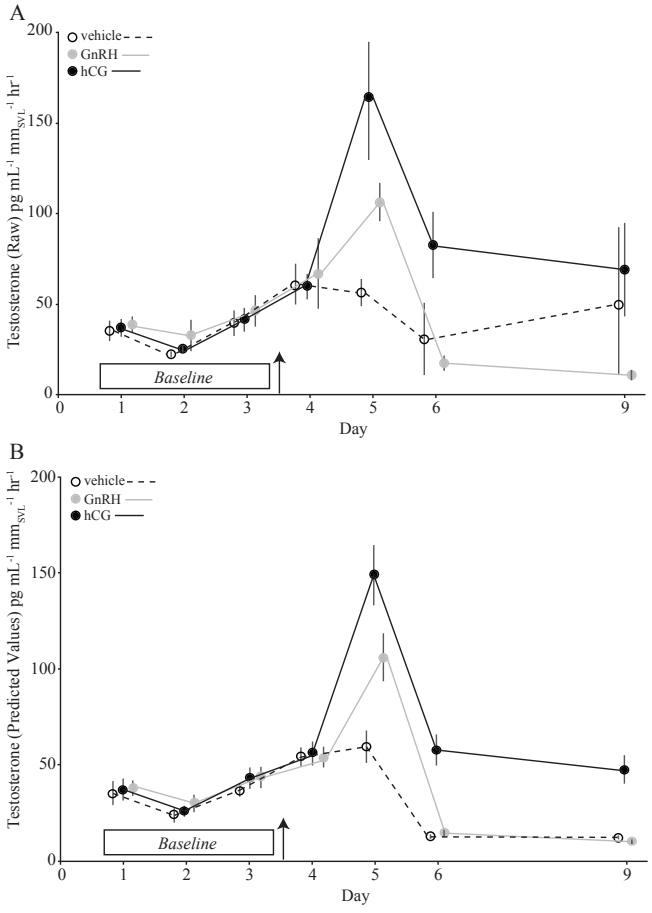


Fig. 3. (a) Raw testosterone values (mean plus or minus 1 SE). Three treatments are jiggered along x-axis for visual clarity. (b) Predicted values (back-transformed) from final model (Day, Treatment, Day*Treatment, and lnCr as covariate). Mean plus or minus 1 S.E. Three treatments are jiggered along x-axis for visual clarity. Arrows on x-axes indicate timing of injection.

Each of the three statistical models identified significant effects of the predictor variables (Table 1). Especially relevant is the interaction between Day and Treatment and the improved model fit with the inclusion of Cr. This improved fit when correcting for Cr concentrations can also be observed in the pattern of variance when comparing the raw T to the adjusted values from Model 3 (Fig. 3).

The post-hoc orthogonal contrasts revealed that the three treatments did not differ from each other during the *Baseline* period (all $p > 0.6$) or for the *Integrated Challenge* period when Days 4 and 5 were combined (all $p > 0.12$). For the *Acute Challenge* period, which was limited to Day 5, there was a significant difference between Saline and hCG groups ($t_{156} = 2.55$, $p = 0.012$) and no difference between GnRH and Saline ($t_{156} = 1.75$, $p = 0.082$) or between GnRH and hCG ($t_{156} = 0.84$, $p = 0.40$). During the *Recovery* period (Days 6–9) there was a significant difference between Saline and hCG groups ($t_{156} = 6.56$, $p < 0.0001$) and between GnRH and hCG ($t_{156} = 6.92$, $p < 0.0001$) and no difference between GnRH and Saline ($t_{156} = 0.43$, $p = 0.67$).

3.2.2. Repeatability

Testosterone exhibited moderate yet significant repeatability for the agreement repeatability estimate (*ICC* (95% CI) = 0.28 (0.11–0.55)) and contained significant within-subject variance (σ_w (95% CI) = 0.11 (0.09–0.13). We found a nominally higher adjusted repeatability estimate when T was corrected for Cr levels (*ICC* (95% CI) = 0.32 (0.14–0.58)), and, more importantly, a reduction in the within-subject variance (σ_w (95% CI) = 0.03 (0.03–0.04).

3.2.3. Repeatedly co-sampling behavior and testosterone

When tested 24 h following injections (i.e. Day 4), frogs in the saline and GnRH treatment groups exhibited low and identical numbers of choices, whereas frogs in the hCG group exhibited more than double the number of choices. In contrast, on Day 9, all three treatment groups exhibited low and identical choice frequencies (Fig. 4).

3.2.4. Sample drying, parallelism and recovery efficacy

We compared two drying methods for Cr estimates using paired split samples (see Methods). Both methods yielded indistinguishable Cr concentrations and the water samples that housed male frogs exhibited Cr levels 3–5 times higher than the control sample that was near the detection limit for the assay (Supplementary figure 8).

Testosterone estimates exhibited clear parallelism in *Phase 2* as demonstrated by a comparison of slopes test (Δ slope = 5.2×10^{-5}, SE = 0.0001, $t_3 = 0.94$, $p = 0.42$; Supplementary figure 9), as did Cr (Δ slope

Table 1

Model	Effects	P-value	−2 Log Likelihood	AICc
1			540.4	552.8
	Day: $F_{6,175} = 8.18$	<0.01		
	Treatment: $F_{2,175} = 8.19$	<0.001		
2			520.7	537.4
	Day: $F_{6,173} = 9.51$	<0.001		
	Treatment: $F_{2,173} = 2.0$	0.14		
	Day × Treatment: $F_{12,173} = 10.43$	<0.0001		
3			356.5	411.2
	Day: $F_{6,156} = 36.0$	<0.0001		
	Treatment: $F_{2,156} = 8.9$	<0.0001		
	Day × Treatment: $F_{12,156} = 5.2$	<0.0001		
	lnCr: $F_{1,156} = 56.1$	<0.0001		

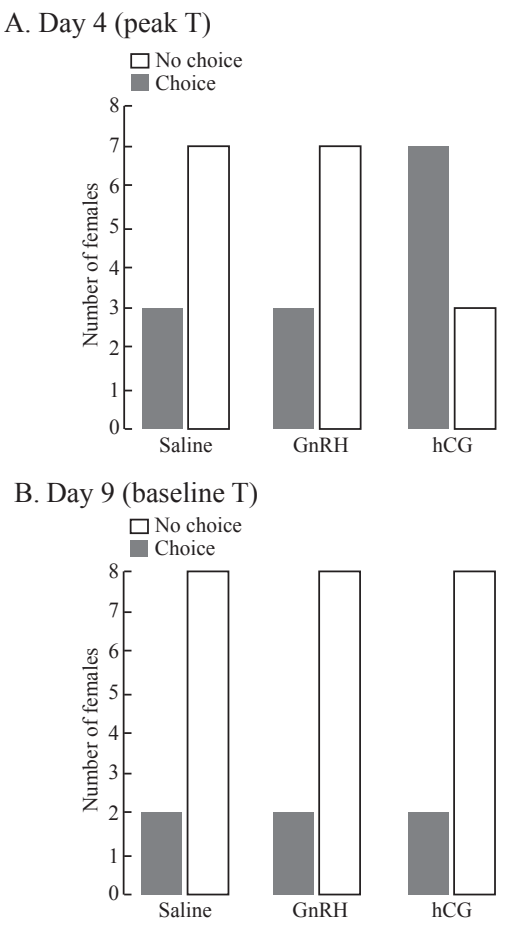


Fig. 4. Phonotaxis results from Day 4 (A) during peak T and Day 9 (B) during recovery-to-baseline T. Choices for either stimuli (whine or whine-chuck) are included under Choice.

= 0.001, SE = 0.18, t_4 = 0.05, p = 0.97; Supplementary figure 10). Stripping was 99.8% effective, and the stripped/unspiked sample registered a very low concentration (3.32 pg mL^{-1}) which we subtracted from the stripped/spiked to calculate recovery. Recovery estimates for T were moderate and invariant (mean, SD) within and among low (40.9, 4.4%), medium (46.1, 1.1%) and high (47.2, 4.3%) concentrations of cold-spiked T (see Supplementary materials S10). Given this consistency, we report the observed hormone concentrations, not attempting to correct for estimated loss during extraction. Intra- and inter-assay CVs, respectively, were 6.2% and 9.7% for all T plates and 1.0% and 4.8% for all Cr plates.

4. Discussion

We established and validated minimally invasive methods for measuring excreted water-borne testosterone that capture precise and physiologically informative estimates. To our knowledge this is the first water-borne androgen validation for an amphibian. We found that a 60-min water bath was sufficient to simultaneously estimate testosterone and creatinine and can be conducted in concert with behavioral testing without interfering with those measurements. Though the pattern in T excretion is clear using raw estimates, by correcting for urine concentration using co-sampled Cr the T estimates were significantly improved. This suggests that urine is a major source of excreted steroids and is consistent with the highly variable (among- and within-individuals) creatinine concentrations across sampling timepoints. It is also worth noting, however, that the current study used potent secretagogues to elevate T, whereas future experimental studies (e.g. population or individual differences) may examine more subtle biological effects where the Cr correction proves even more important at improving signal-to-noise ratios. Therefore, if feasible, we recommend future studies co-sample creatinine to adjust hormones estimates. Though significantly correlated, measuring Cr in lieu of T is probably not sufficiently precise, given the variance in the strength of these correlations.

In males, an hCG challenge induced a large elevation in T within 48 h of injection while vehicle injections had no effect, and GnRH challenges may have induced an intermediate elevation. Elevated T then persisted for at least 24 h before recovering to baseline. These predictable temporal dynamics and sex differences help validate the biologically informative nature of these methods and serve as a good alternative to tissue harvesting for measuring circulating steroid concentrations, which is often lethal in small amphibians such as túngara frogs. The observation that hCG induced a higher peak T than GnRH, which also appears to persist longer, is challenging to interpret because this may simply be a dose-dependent pattern with a resulting lag in clearance rates. That said, there does appear to be an indication of negative feedback driving T levels to baseline or possibly even sub-baseline levels after recovery in the vehicle and GnRH groups. A sperm-release study by Silla et al. (2019) demonstrated that across a wide range of dosages, hCG was more effective at inducing spermiation compared to GnRH, though a lower hCG dose (40 IU g^{-1}) yielded maximum sperm-release. Therefore, the differential effects observed here—in both excreted T and phonotaxis—may reflect a higher secretory capacity of hCG compared to GnRH. The Cr-corrected levels of T may indicate that vehicle injected frogs have slightly elevated T after several days of treatment, which then decline to sub-baseline concentrations during the recovery period. This pattern could be caused by the natural stimulation of the HPG axis in these control males through repeated daily exposure to standing water, which in the field and lab induces the reliable onset of sexual behavior, including vocalizing and phonotaxis (Baugh, pers. observ.). Hence, the utility of incorporating a control group into a study design such that this effect can be accounted for.

Baseline concentrations of T were repeatable: approximately thirty percent of the variation was explained by individual differences and the remainder represents within-individual variance. This suggests that (1) the procedure has sufficient fidelity and sensitivity to capture biologically relevant variation; (2) males likely vary in their underlying HPG axis activity under baseline conditions; and (3) such individual variation argues for the deployment of within-subject designs for future studies, which are feasible using these non-invasive methods. The nominal increase in the *ICC* estimate and the larger reduction of the within-subject variance in the adjusted repeatability analysis is further evidence indicated that correcting excreted hormone concentrations using Cr improves the signal-to-noise ratiothe sign.

We assume that the timeline for T elevation in plasma is much faster following pharmacological challenge compared to the excreted water-borne signal. Consistent with this assumption, we found that T spiked in the hCG group a day after the peak phonotaxis behavior in this group. One interpretation of this result is that elevated circulating T increases male phonotaxis after it exceeds a particular threshold, which presumably was not met in the GnRH and vehicle treatment groups. It is also possible, however, that hCG elevated several different steroids that induce sexual behavior in males (as hCG does in female túngara frogs; Baugh et al., 2018). More research is needed along these lines. At minimum, the behavioral findings suggest that steroid concentrations may underlie sexually proceptive behavior such as male phonotaxis. Phonotaxis testing in amphibians is almost exclusively conducted in reproductively ready females, whereas sexual behavior in males is often evaluated by measuring vocal behavior (Gerhardt and Huber, 2002) including in túngara frogs (Bernal et al. 2009). In several species, however, males and even juveniles of both sexes have been shown to exhibit species-typical phonotaxis (Lea et al., 2000; Baugh et al., 2012; Pfennig et al., 2013). Therefore, phonotaxis testing offers the opportunity to directly compare the sexes without the confounding effect of task

differences and therefore understanding the physiological basis for behavior including sex differences therein (Baugh and Ryan, 2017). For example, both empirical and theoretical evidence strongly indicates that in many species females exhibit greater behavioral sensitivity and discrimination in response to male sexual signals (Kirkpatrick and Ryan, 1991; Andersson, 1994). Although we do not yet understand the physiological basis for this evolutionary critical sex difference, it is possible that one contributing factor is the hormonal milieu—an example of this is the sexually divergent effect that the neuropeptide arginine vasotocin has on phonotaxis in túngara frogs (Baugh and Ryan, 2017). The finding here that elevated T in males coincides with heightened phonotaxis, coupled with earlier studies indicating that estradiol and progesterone (also elevated by hCG) have this effect in female frogs (Chakraborty and Burmeister, 2009; Lynch and Wilczynski, 2005; Lynch et al., 2006; Bastien et al., 2018; Baugh et al., 2018, 2019; Gall et al., 2019), may suggest non-overlapping hormonal bases for phonotaxis between the sexes. Androgens, however, have been shown to underlie sexual behavior in both sexes in fish (Gabor and Grober, 2010) and neural receptors are present but sexually dimorphic in túngara frogs (Chakraborty and Burmeister, 2010). More research is needed to understand these endocrine mechanisms, including pharmacological control of steroid conversion (e.g. aromatization of testosterone to estrogen; Chakraborty and Burmeister, 2009).

In summary, we demonstrated that a 60-min water bath with a correction for Cr represents a promising and non-invasive methodology for capturing biologically informative T concentrations while permitting the experimenter to repeatedly sample multiple hormones and behavior in individual amphibians. Using these methods we demonstrated several empirical results. First, males have higher T than females, which were also significantly repeatable at the individual level. Second, frogs exhibit a substantial increase in T within 48 h of challenge (hCG, GnRH). Third, the frequency of phonotaxis behavior was positively associated with pharmacologically elevated T. And lastly, adjusting T levels using simultaneous measures of Cr improves the accuracy of the T estimates, which is likely due to highly variable urine concentrations in the water samples, thereby indicating that urine likely serves as the major source of excreted steroids. Because of the simplicity and non-invasive nature of the sample collection, this set of methods can be deployed in field or lab settings, will be relevant for conservation physiology, and can be conducted simultaneously or sequentially with behavioral measures.

CRediT authorship contribution statement

Alexander T. Baugh: Conceptualization, Methodology, Formal analysis, Investigation, Supervision, Funding acquisition, Writing - original draft. **Sophie L. Gray-Gaillard:** Conceptualization, Methodology, Formal analysis, Investigation, Writing - review & editing.

Acknowledgements

We thank S.Silver, N. Lee, J. Woo, N. Ambiel and D. Lee for their assistance with data collection, three anonymous reviewers for improvements to the manuscript and the Swarthmore College Research Fund for providing funding. The Institutional Animal Care and Use Committee at Swarthmore College approved the protocols used in this study.

Appendix A. Supplementary data

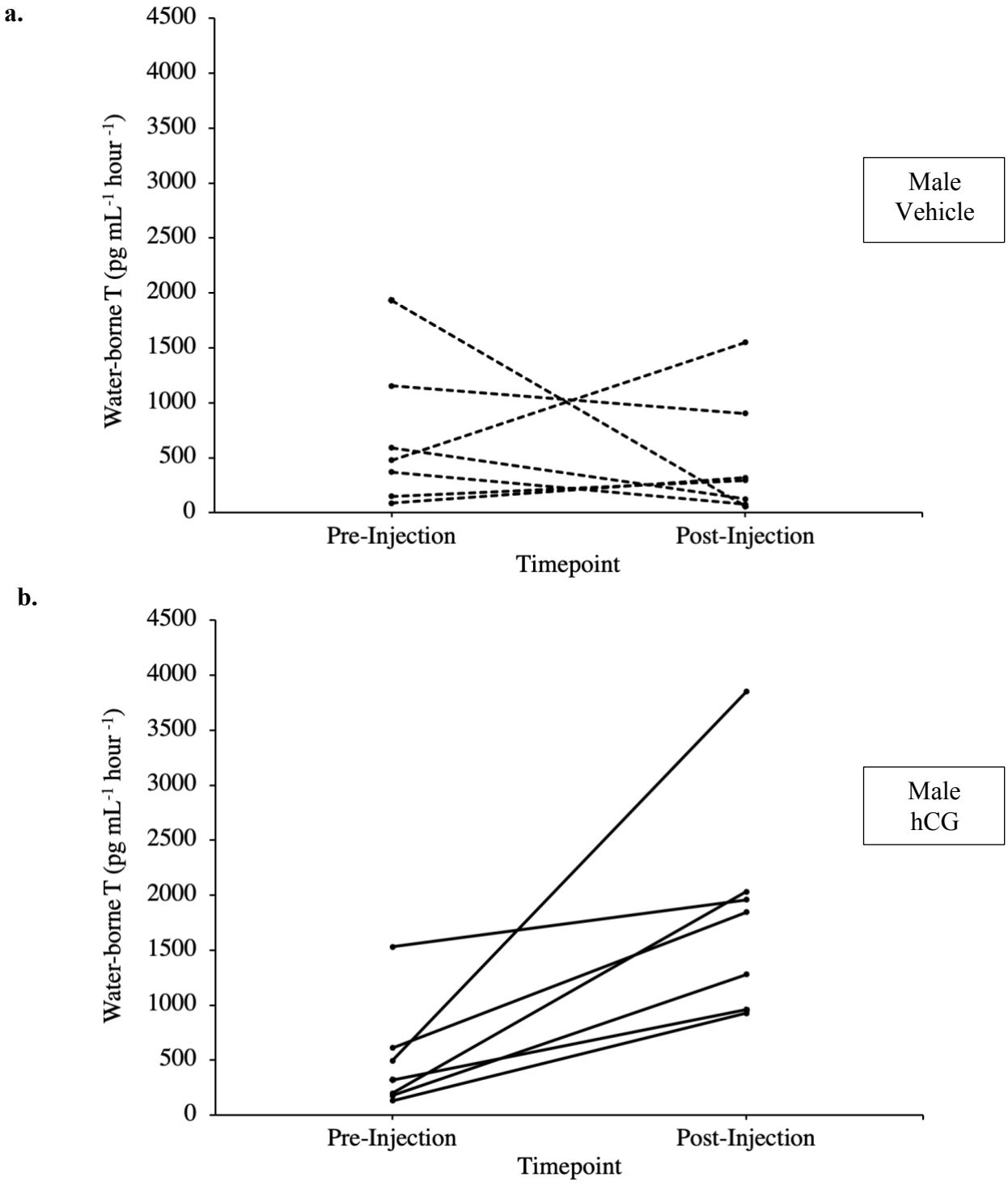
Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygcen.2020.113638.

References

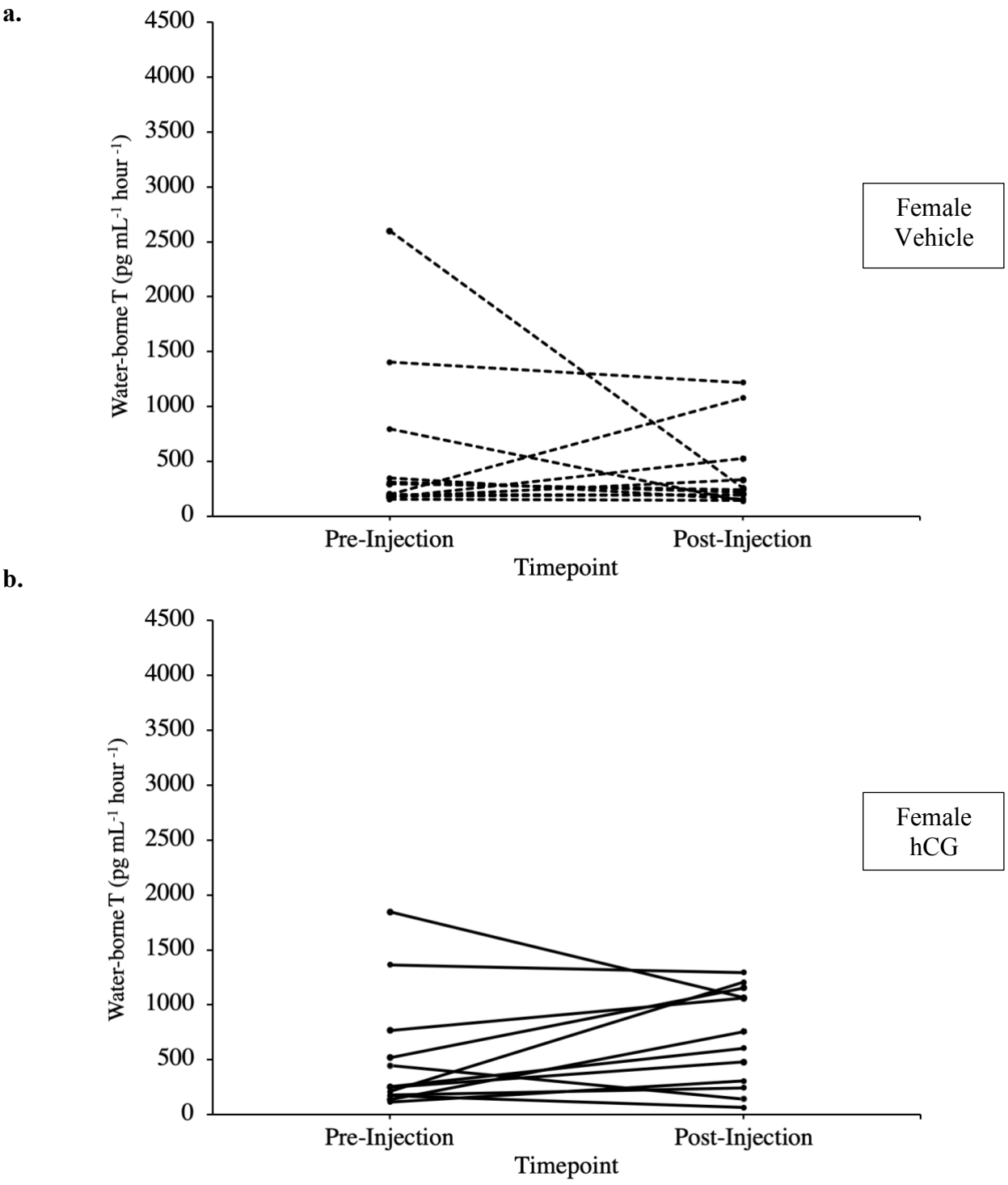
Andersson, M., 1994. Sexual Selection. Princeton University Press.
Baugh, A.T., Ryan, M.J., 2010a. Temporal updating during phonotaxis in male túngara frogs (*Physalaemus pustulosus*). Amphibia-Reptilia 31, 449–454.
Baugh, A.T., Ryan, M.J., 2010b. The development of sexual behaviour in túngara frogs. J. Comp. Psychol. 124, 66–80.
Baugh, A.T., Hau, M., Naguib, M., van Oers, K., 2013. Initial reactivity and magnitude of the acute stress response linked to exploratory behavior in great tits (*Parus major*). Gen. Comp. Endocrinol. 189, 96–104.
Baugh, A.T., Hoke, K.L., Ryan, M.J., 2012. Development of communication behaviour: receiver ontogeny in túngara frogs and a prospectus for a behavioural evolutionary development. Sci. World J. 2012, 1–10.
Baugh, A.T., Ryan, M.J., 2017. Vasotocin induces sexually dimorphic effects of acoustically-guided behavior in a tropical frog. J. Comp. Physiol. A. 203, 265–273.
Baugh, A.T., Witonsky, K.R., Davidson, S.C., Hyder, L., Hau, M., van Oers, K., 2017a. Novelty induces behavioral and glucocorticoid responses in a songbird (*Parus major*) artificially selected for divergent personalities. Anim. Behav. 130, 221–231.
Baugh, A.T., Senft, R.A., Firke, M., Lauder, A., Schroeder, J., Meddle, S.L., van Oers, K., Hau, M., 2017b. Risk-averse personalities have a systemically potentiated neuroendocrine stress axis: a multilevel experiment in *Parus major*. Horm. Behav. 93, 99–108.
Bastien, B., Farley, G., Ge, F., Malin, J.S., Simon-Plumb, C.L., Pulley, D.M., Yang, C., Baugh, A.T., 2018. The waiting-mating game: condition dependent mate sampling in female grey treefrogs (*Hyla versicolor*). Front. Ecol. Evol. 6, 140.
Baugh, A.T., Bastien, B., Still, M., Stowell, N., 2018. Validation of water-borne steroid hormones in tropical frog (*Physalaemus pustulosus*). Gen. Comp. Endocrinol. 261, 67–80.
Baugh, A.T., Bee, M., Gall, M., 2019. The paradox of hearing at the lek: auditory sensitivity increases after breeding in female gray treefrogs *(Hyla chrysoscelis*). J. Comp. Physiol. A. 205, 629–639.
Bernal, X.E., Akre, K.L., Baugh, A.T., Rand, A.S., Ryan, M.J., 2009. Female and male behavioral response to advertisement calls of graded complexity in túngara frogs. Behav. Ecol. Sociobiol. 63, 1269–1279.
Bosch, J., Márquez, R., 2002. Female preference function related to precedence effect in an amphibian 655 anuran (Alytes cisternasii): tests with non-overlapping calls. Behav. Ecol. 13, 149–153.
Breuner, C.W., Wingfield, J.C., 2000. Rapid behavioral response to corticosterone varies with photoperiod and dose. Horm. Behav. 37, 23–30.
Chakraborty, M., Burmeister, S.S., 2009. Estradiol induces sexual behavior in female túngara frogs. Horm. Behav. 55, 106–112.
Chakraborty, M., Burmeister, S.S., 2010. Sexually dimorphic androgen and estrogen receptor mRNA expression in the brain of túngara frogs. Horm. Behav. 58, 619–627.
Cheng, M.F., 1992. For whom does the female dove coo? A case for the role of self-stimulation. Anim. Behav. 43, 1035–1044.
Delehanty, B., Hossain, S., Jen, C.C., Crawshaw, G.J., Boonstra, R., 2015. Measurement of free glucocorticoids: quantifying corticosteroid-binding globulin binding affinity and its variation within and among mammalian species. Conserv. Physiol. 3, 1–13.
Dingemanse, N.J., Dochtermann, N.A., 2013. Quantifying individual variation in behaviour: mixed-efect modelling approaches. J. Anim. Ecol. 82, 39–54.
Gabor, C.R., Grober, M.S., 2010. A potential role of male and female androgen species recognition in a unisexual-bisexual mating complex. Horm. Behav. 57, 427–433.
Gabor, C.R., Fisher, M.C., Bosch, J., 2013a. A non-invasive stress assay shows that tadpole populations infected with *Batrachochytrium dendrobatidis* have elevated corticosterone levels. PLoS One 8, e56054.
Dingemanse, N.J., Edelaar, P., Kempenaers, B., 2010. Why is there variation in baseline glucocorticoid levels? Trends Ecol. Evol. 25, 261–262.
Gabor, C., Bosch, J., Fries, J., Davis, D., 2013b. A noninvasive water-borne hormone assay for amphibians. Amphibia-Reptilia. 34, 151–162.
Gabor, C.R., Fisher, M.C., Bosch, J., 2015. Elevated corticosterone levels and changes in amphibian behavior are associated with *Batrachochytrium dendrobatidis* (Bd) infection and Bd lineage. PLoS One 10, e0122685.
Gall, M., Bee, M., Baugh, A.T., 2019. The difference a day makes: breeding remodels hearing, hormones and behavior in female Cope's gray treefrogs *(Hyla chrysoscelis*). Horm. Behav. 108, 62–72.
Gerhardt, H.C., Huber, F., 2002. Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions. University of Chicago Press.
Good, T., Khan, M.Z., Lynch, J.W., 2003. Biochemical and physiological validation of a corticosteroid radioimmunoassay for plasma and fecal samples in oldfield mice (*Peromyscus polionotus*). Physiol. Behav. 80, 405–411.
Hau, M., Casagrande, S., Ouyang, J.Q., Baugh, A.T., 2016. Glucocorticoid-Mediated Phenotypes in Vertebrates: Multilevel Variation and Evolution. In: Naguib, M., Mitani, J.C., Simmons, L.W., Barrett, L., Healy, S., Zuk, M. (Eds.), Advances in the Study of Behavior. Academic Press, pp. 41–115.
Kindermann, C., Narayan, E.J., Hero, J.M., 2012. Urinary corticosterone metabolites and chytridiomycosis disease prevalence in a free-living population of male Stony Creek frogs (Litoria wilcoxii). Comp. Biochem. Physiol. Part A Molec. Integ. Physiol. 162, 171–176.
Kirkpatrick, M., Ryan, M.J., 1991. The paradox of the lek and the evolution of mating preferences. Nature 350, 33–38.
Lea, J., Halliday, T., Dyson, M., 2000. Reproductive stage and history affect the phonotactic preferences of female midwife toads, *Alytes muletensis*. Anim. Behav. 60, 423–427.
Leary, C., Baugh, A.T., 2020. Glucocorticoids, male sexual signals, and mate choice in females: implications for sexual selection. Gen. Comp. Endocrinol. 288, 113354.
Lee, C., Goeger, D.E., 1998. Interference of 6b-Hydroxycortisol in the quantification of urinary free cortisol by immunoassay and its elimination by solid-phase extraction. Clin. Biochem. 31, 229–233.

Lynch, K.S., Crews, D., Ryan, M.J., Wilczynski, W., 2006. Hormonal state influences aspects of female mate choice in the túngara frog (*Physalaemus pustulosus*). Horm. Behav. 49, 450–457.
Lynch, K.S., Wilczynski, W., 2005. Gonadal steroids vary with reproductive stage in a tropically breeding female anuran. Gen. Comp. Endocrinol. 143, 51–56.
Lynch, K.S., Wilczynski, W., 2006. Social regulation of plasma estradiol concentration in a female anuran. Horm. Behav. 50, 101–106.
Marler, C.A., Ryan, M.J., 1996. Energetic constraints and steroid hormone correlates of male calling behaviour in the túngara frog. J. Zool. Lond. 240, 397–409.
Narayan, E.J., 2013. Non-invasive reproductive and stress endocrinology in amphibian conservation physiology. Conserv. Physiol. 1, 2013.
Narayan, E.J., Forsburg, Z.R., Davis, D.R., Gabor, C.R., 2019. Non-invasive Methods for Measuring and Monitoring Stress Physiology in Imperiled Amphibians. Front. Ecol. Evol. 7, 431. https://doi.org/10.3389/fevo.2019.00431.
Narayan, E., Molinia, F., Christi, K., Morley, C., Cockrem, J., 2010. Urinary corticosterone metabolite responses to capture, and annual patterns of urinary corticosterone in wild and captive endangered Fijian ground frogs (*Platymantis vitiana*). Aust. J. Zool. 58, 189–197.
Newman, A.E.M., Chin, E.H., Schmidt, K.L., Bond, L., Wynne-Edwards, K.E., Soma, K.K., 2008. Analysis of steroids in songbird plasma and brain by coupling solid phase extraction to radioimmunoassay. Gen. Comp. Endocrinol. 155, 503–510.
Odum, R.A., Zippel, K.C., 2008. Amphibian water quality: approaches to an essential environmental parameter. Intern. Zoo Yearbook 42, 40–52.
Pfennig, K.S., Rodriguez Moncalvo, V.G., Burmeister, S.S., 2013. Diet alters ontogeny of species recognition in juvenile toads. Biol. Lett. 9, 20130599.
Rand, A.S., Ryan, M.J., Wilczynski, W., 1992. Signal redundancy and receiver permissiveness in acoustic mate recognition by the túngara frog Physalaemus pustulosus. American Zoologist 32, 81–90.
Romero, L.M., Reed, J.M., 2005. Collecting baseline corticosterone samples in the field: is under three minutes good enough? Comp. Biochem. Physiol. A 140, 73–79.
Ryan, M.J., 1985. The Túngara Frog: A Study in Sexual Selection and Communication. University of Chicago Press, Chicago.
Ryan, M.J., Akre, K.A., Baugh, A.T., Bernal, X., Lea, A., Leslie, C., Still, M., Rand, A.S., 2019. Consistently positive and strong sexual selection: 19 years of female mate choice preferences. Am. Nat. 194, 125–134.
Ryan, M.J., Rand, A.S., 1990. The sensory basis of sexual selection for complex calls in the túngara frog Physalaemus pustulosus sexual selection for sensory exploitation. Evolution 44, 305–314.
Ryan, M.J., Rand, A.S., 2003. Sexual selection and female preference space: how female túngara frogs perceive and respond to complex population variation in acoustic mating signals. Evolution 57, 2608–2618.
Ryan, M.J., Rand, W., Hurd, P.L., Phelps, S.M., Rand, A.S., 2003. Generalization in response to mate recognition signals. Am. Nat. 161, 380–394.
Silla, A.J., McFadden, M.S., Byrne, P., 2019. Hormone-induced sperm-release in the critically endangered Booroolong frog (Litoria booroolongensis): effects of gonadotropin-releasing hormone and human chorionic gonadotropin. Conserv. Physiol. 2019; 7(1):coy080.
Still, M.B., Lea, A.M., Hofmann, H.A., Ryan, M.J., 2019. Multimodal stimuli regulate reproductive behavior and physiology in male túngara frogs. Horm. Behav. 115, 104546.
Touma, C., Palme, R., 2005. Measuring fecal glucocorticoid metabolites in mammals and birds: the importance of validation. Ann. N.Y. Acad. Sci. 1046, 54–74.
Trainor, B.C., Lin, S., Finy, M.S., Rowland, M.R., Nelson, R.J., 2007. Photoperiod reverses the effects of estrogens on male aggression via genomic and non-genomic pathways. Proc. Nat. Acad. Sci. USA 104, 9840–9845.
Trainor, B.C., Finy, M.S., Nelson, R.J., 2008. Rapid effects of estradiol on male aggression depend on photoperiod in reproductively non-responsive mice. Horm. Behav. 53, 192–199.
Vu, M., Weiler, B., Trudeau, V.L., 2017. Time- and dose-related effects of a gonadotropin-releasing hormone agonist and dopamine antagonist on reproduction in the Northern leopard frog (*Lithobates pipiens*). Gen. Comp. Endocrinol. 254, 86–96.

S1. Depiction of water bath apparatus. A túngara frog in 25 mL of ‘frog water’ bath apparatus consisting of a translucent Dixie cup with another Dixie cup secured above, allowing the frog breathing headspace and ensuring the frogs were immersed in the bath throughout the entire collection period. Water baths were constituted in 30 L of RO water with solutes dissolved: 1.2 g $CaCl_2$, 1.38 g $MgSO_4$, 1.08 g $KHCO_3$, 0.9 g $NaHCO_3$ and 0.038 g of a commercial trace-element.



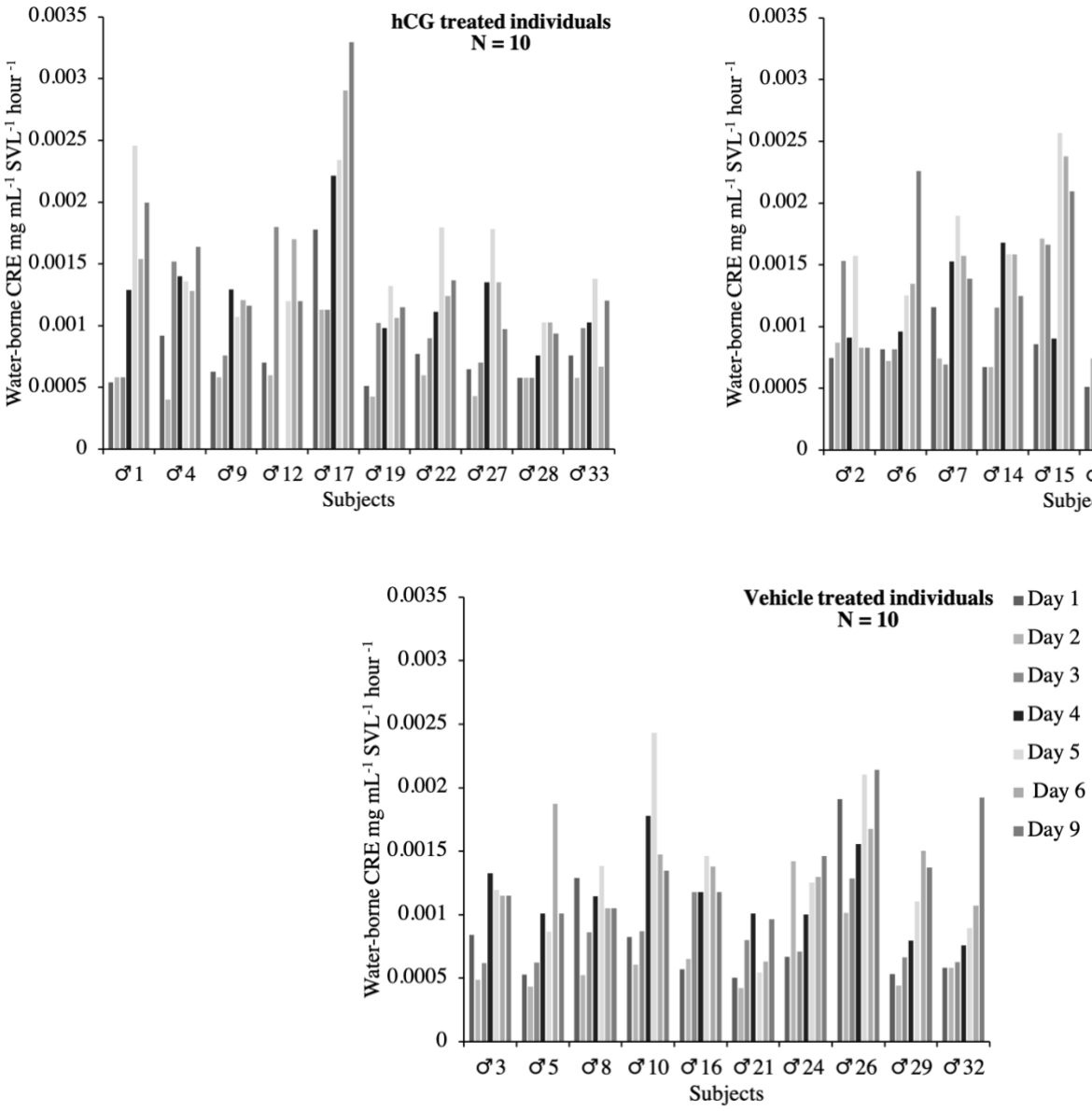
S2. Phase I: Individual water-borne T concentrations of male water samples taken at two timepoints 24 hours apart. Lines indicate the within-subject concentration changes across timepoint. (a) Individual T concentrations for vehicle treatment (saline, $N_{vehicle} = 8$), and (b) individuals T concentrations for hCG (500 IU g-1, $N_{hCG} = 8$) treatment.



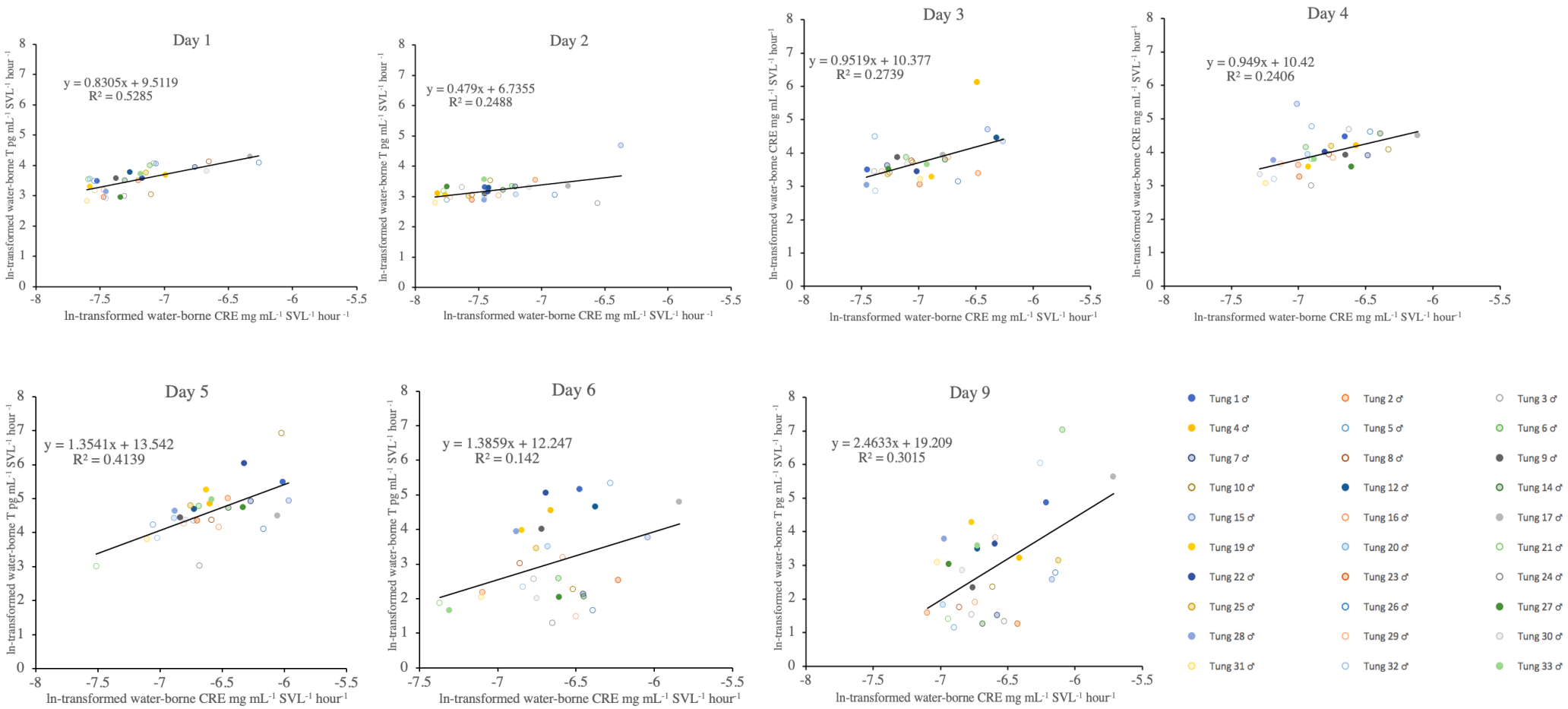
S3. Phase I: Individual water-borne T concentrations of female water samples taken at two timepoints 24 hours apart. Lines indicate the within-subject concentration changes across timepoint. (a) Individual T concentrations for vehicle treatment (saline, $N_{vehicle} = 13$), and (b) individuals T concentrations for hCG (500 IU g-1, $N_{hCG} = 13$) treatment.

Percentage of pooled sample in assay buffer (%)	**Concentration (pg mL^{-1}) determined by testosterone ELISA** ± SD
100	773.61 ± 137.30
50	294.42 ± 19.73
25	99.461 ± 38.62
12.5	33.279 ± 0 (run in singlet)
6.25	out of range
3.0125	out of range
1.5625	out of range

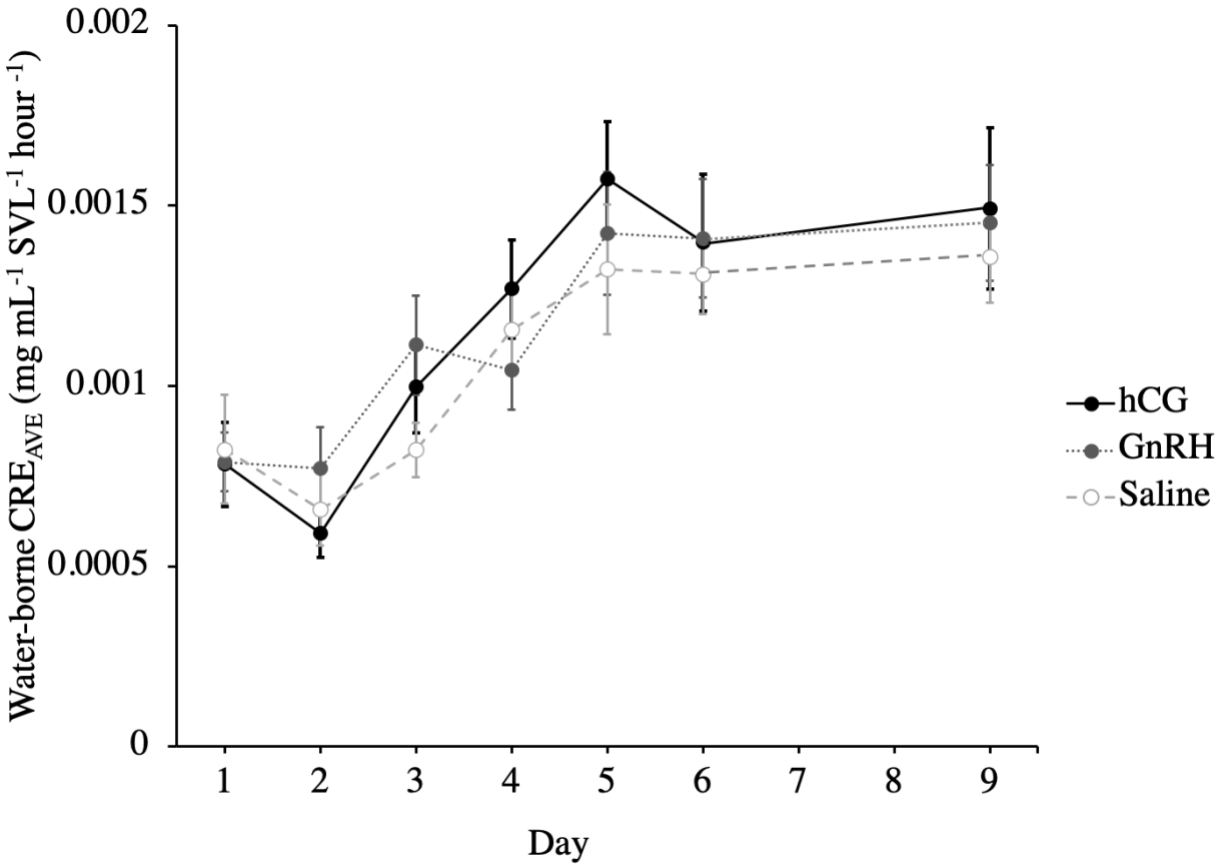
S4. Phase I: Water-borne T concentration results from a preliminary parallelism study derived from a pooled sample of all male water samples including both timepoints 24 hours apart (preinjection and post-injection) and treatment groups: hCG (500 IU g^{-1}) and vehicle (saline). Pooled sampled were run in duplicate (unless noted otherwise) at a dilution factor of 1:1 (250 uL reconstitution volume for a 25 mL water bath sample).



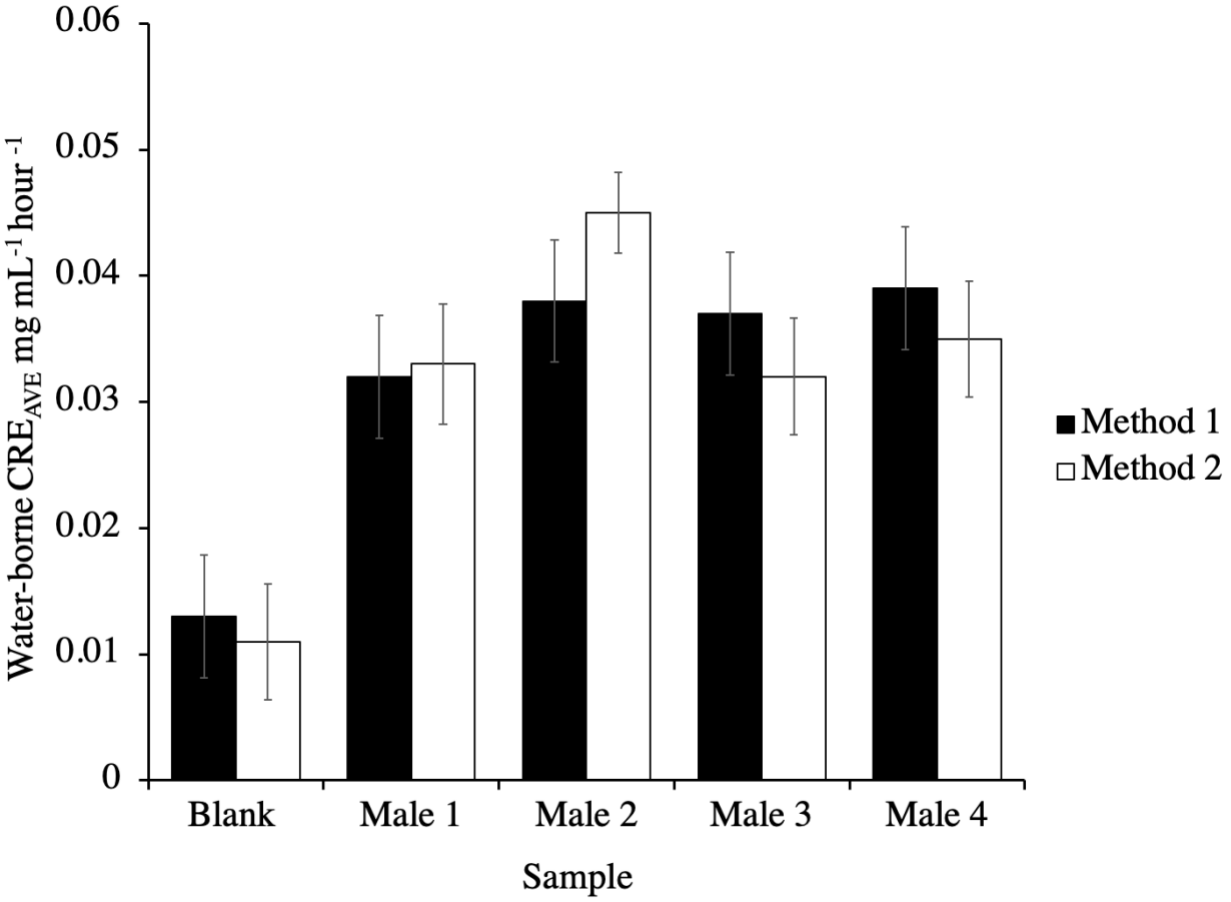
S5. Phase II: Water-borne Cr concentrations are depicted for each subject on every day of sample collection (Days 1-6 and Day 9). These graphs show varied Cr concentrations within and among-individuals. Graphs are separated per treatment groups: (a) hCG (500 IU g^{-1}, N = 10), (b) GnRH (0.5 ug g^{-1}, N = 10), and (c) vehicle (saline, N = 10) injected males.



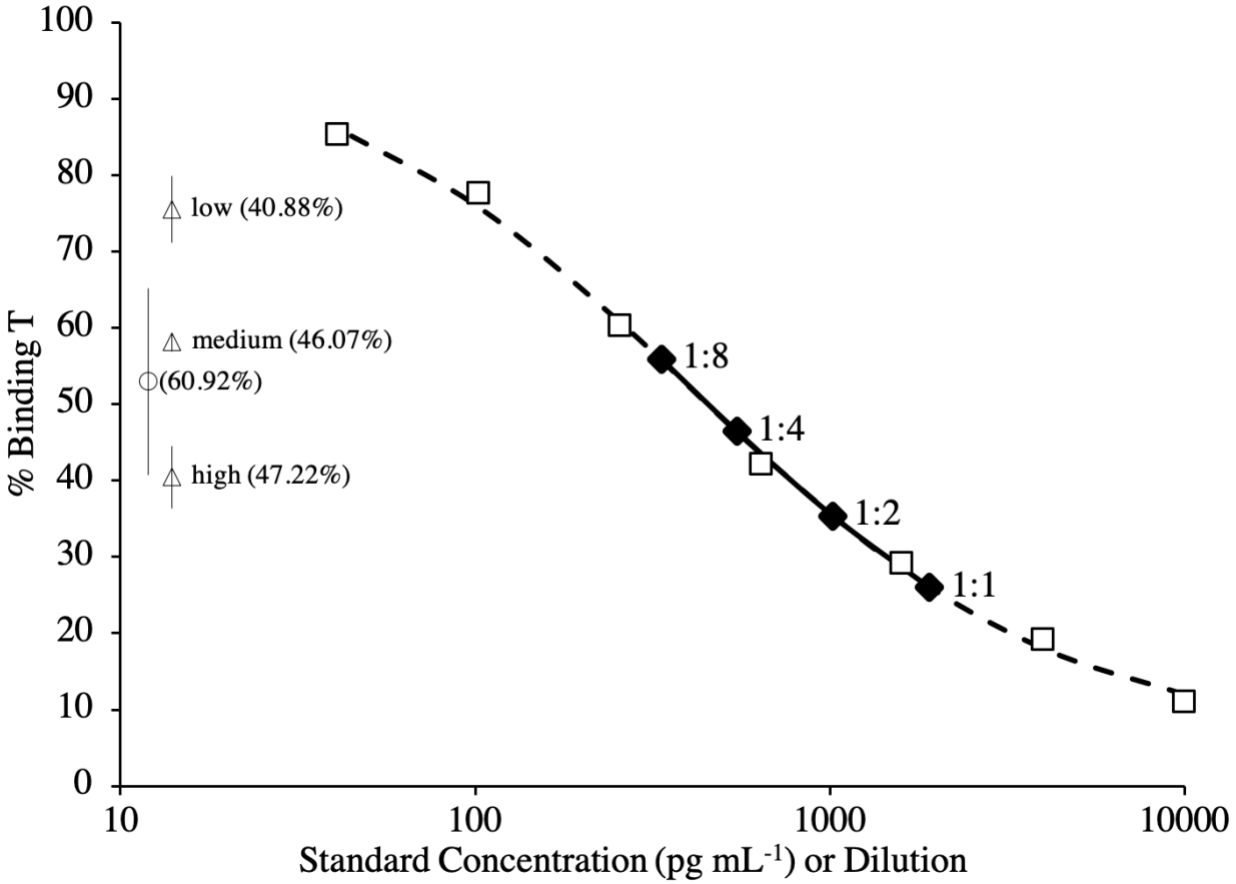
S6. Phase 2. Linear regressions of ln-transformed water-borne Cr and ln-transformed water-borne T across (a-f) Days 1-6 and (g) Day 9 for all individuals. Solid circles represent hCG (500 IU g^{-1}, N = 10), lightly shaded circles GnRH (0.5 ug g^{-1}, N = 10), and open circles vehicle (saline, N = 10) injected males. These results demonstrated a significantly positive between Cr and T (Linear regression: r^2 = range from 0.142 - 0.529, N = 30, F = 4.63 - 31.39; $p < 0.04$ for all linear regressions).



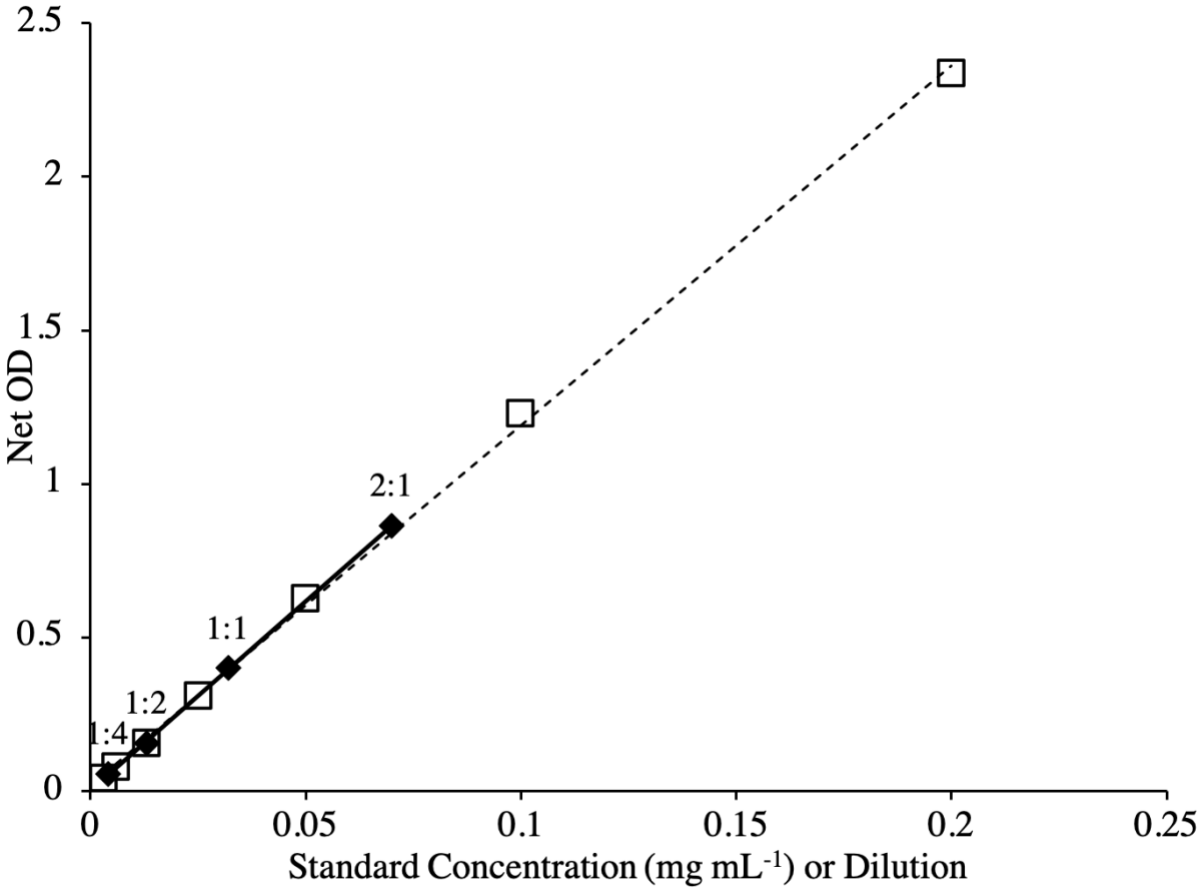
S7. Phase II: Water-borne Cr concentrations (mean ± 1 SE) for hCG (500 IU g^{-1}, N = 10), GnRH (0.5 ug g^{-1}, N = 10), and vehicle (saline, N = 10) injected males. Water samples were taken on Days 1-6 and Day 9, and concentrations were determined by creatinine EIA. Lines indicate average concentration changes for each group across Day.



S8: Phase II: Water-borne CRE (mean ± 1 SE) of split-paired male water bath samples (12 mL, N =4 males and 1 blank 'frog water' sample) to validate two drying methods. Method 1: drying 3 mL of a sample to completion in succession totaling four rounds, and resulting dried extracts were used for CRE detection. Method 2: drying four separate 3 mL aliquots of a sample and pooling the reconstituted samples prior to plating. The reported CRE concentrations between drying methods for each sample showed negligible differences



S9. Phase II: Parallelism for water-borne T using the DetectX® Testosterone ELISA Kit (Arbor Assays, Ann Arbor, MI; Cat. No. k032). Dashed lines and open squares illustrate the standard curve (four parameter equation; Softmax Pro), and solid line and solid diamonds indicate a pooled sample at four dilutions. Based on parallelism, we processed samples at 1:1 dilution factor (250 uL reconsitution volume for a complete 25 mL water bath sample, or 125 uL reconstitution volume for a halved (12.5 mL) water bath subsample). Open triangles represent mean percent binding ± 1 SD for high (1,600 pg mL^{-1}), medium (640 pg mL^{-1}), and low (256 pg mL^{-1}) concentration cold-spiked recovery samples with percent recovery shown in parentheses. Open circle represents mean percent binding ± 1 SD for cold-spike recovery sample (pooled water bath sample stripped and cold-spiked to 640 pg mL^{-1} using purified T prior to extraction), and percent recovery is shown in parentheses. The standard curve and dilution curve were parallel demonstrated by comparison of slopes (Δ slope = 5.2 x 10^{-5}, SE = 0.0001, t_3 = 0.94, p = 0.42).



S10. Phase II: Parallelism for water-borne CRE using the Urinary Creatinine Detection Kit (DetectX® kit, Arbor Assays, Ann Arbor, MI; Cat. No. K002). Dashed lines and open squares illustrate the standard curve (four parameter equation; Softmax Pro), and solid line and solid diamonds indicate a pooled sample at four dilutions. Based on parallelism, we processed samples at 1:1 dilution factor. Serial dilutions were parallel to the standard curve as demonstrated by a comparisons of slopes test (delta slope = 0.001, SE = 0.18, $t_4 = 0.05$, $p = 0.97$).