INTERCALIBRATION EXERCISE USING A STICKLEBACK ENDOCRINE DISRUPTER SCREENING ASSAY

Yvonne T. Allen,*† Ioanna Katsiadaki,‡ Tom G. Pottinger,§ Cecile Jolly,‖ Peter Matthiessen,§
Ian Mayer,‖ Andy Smith,† Alexander P. Scott,‡ Paul Eccles,‡ Matthew B. Sanders,‡
Kim G.T. Pulman,§ and Stephen Feist‡

†Centre for Environment, Fisheries and Aquaculture Science, Burnham Laboratory, Remembrance Avenue, Burnham-on-Crouch,
Essex, CM0 8HA, United Kingdom
‡Centre for Environment, Fisheries and Aquaculture Science, Weymouth Laboratory, The Nothe, Barrack Road,
Weymouth, Dorset, DT4 9UB, United Kingdom
§Centre for Ecology and Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, LA1 4AP, United Kingdom
‖Department of Biology, University of Bergen, N-5020 Bergen, Norway

Abstract—The Organisation for Economic Cooperation and Development (OECD) is currently validating a short-term fish screening protocol for endocrine disrupters (estrogens, androgens, and their antagonists and aromatase inhibitors), using three core species: fathead minnow, Japanese medaka, and zebrafish. The main endpoints proposed for the first phase of validation of the screen are vitellogenin (VTG) concentration, gross morphology (secondary sexual characteristics and gonado-somatic index), and gonadal histopathology. A similar protocol is concurrently being developed in the United Kingdom using the three-spined stickleback, with identical endpoints to those for the core species and, in addition, a unique androgen-specific endpoint in the form of spiggin (glue protein) induction. To assess the suitability of this species for inclusion in the OECD protocol alongside the core species, an intercalibration was conducted using 17β-estradiol (a natural estrogen) and trenbolone (a synthetic androgen), thus mimicking a previous intercalibration with the core species. All three participating laboratories detected statistically significant increases in VTG in males after 14 d exposure to nominal concentrations of 100 ng/L 17β-estradiol and statistically significant increases in spiggin in females after 14 d exposure to nominal concentrations of 5,000 ng/L trenbolone. The stickleback screen is reliable, possessing both relevant and reproducible endpoints for the detection of potent estrogens and androgens. Further work is underway to assess the relevance and suitability of the screen for weakly acting estrogens, anti-androgens, and aromatase inhibitors.

Keywords—Stickleback   Intercalibration   Endocrine disrupter   Fish screening assay

INTRODUCTION

A working group on endocrine disrupter testing and assessment, formed jointly by the Organisation for Economic Cooperation and Development (OECD) Risk Assessment Advisory Board and the National Coordinators of the OECD Test Guidelines Program, was tasked with identifying and prioritizing new and enhanced test guidelines and developing a testing strategy for endocrine disrupters (Organisation for Economic Cooperation and Development. 1999. Report from the OECD Expert Consultation on Testing in Fish, London, October 28–29, 1998, Test Guidelines Program, Paris, France). This objective reflected the views of OECD member countries that no existing OECD test guidelines were currently appropriate for the detection of endocrine disrupting chemicals (EDCs) in either mammalian or nonmammalian species. There was also a desire that the approaches and procedures used for endocrine disrupter testing should not substantially differ between countries. The endocrine disrupter testing and assessment subsequently developed a tiered, flexible framework for the screening and testing of EDCs.

Several potential screening and testing assays, based upon previously developed designs [1–4] were identified at an expert consultation on fish-testing needs, which needed formal validation. To address this, the OECD Secretariat established a validation management group for ecotoxicological test methods for endocrine disrupters (VMG-Eco [toxicology]) with the remit to oversee the work on validation of EDC fish tests as well as tests using other taxa such as amphibians, birds, and certain invertebrates.

A 21 d fish-screening assay is being taken forward by the VMG-Eco for development and subsequent validation. This protocol [5] currently includes three test species—fathead minnow (*Pimephales promelas*), Japanese medaka (*Oryzias latipes*), and zebra-fish (*Danio rerio*). Essentially, the core endpoints used in this screen are vitellogenin (VTG) concentration, gross morphology (including secondary sexual characteristics where relevant and gonado-somatic index (GSI; gonad weight/body weight·100) and gonadal histopathology. Two phases (termed Phase 1 [A and B] and Phase 2) of validation have already been performed with the core species to assess relevance, sensitivity and reproducibility of the chosen endpoints for known potent and weakly acting EDCs and negative substances. Test chemicals have included potent estrogens and androgens (17β-estradiol [E2] and trenbolone) in Phase 1A [5]; a weak estrogen (4-*tert*-pentylphenol), an anti-androgen (flutamide), and an aromatase inhibitor (prochloraz) in Phase 1B [6]; and potassium permanganate and *n*-octanol (presumed negatives) in Phase 2.

Two major disadvantages have been identified for the three chosen small fish species. First, none are indigenous to United

Table 1. Experimental design used for the stickleback intercalibration[a]

	No. of fish sampled		
	Day 0	Day 14	Day 21
17β-estradiol			
Negative control	8F + 8M	10F + 10M	10F + 10M
Solvent control (0.0001% methanol)		10F + 10M	10F + 10M
10		10F + 10M	10F + 10M
32		10F + 10M	10F + 10M
100		10F + 10M	10F + 10M
Trenbolone			
Negative control	8F + 8M	10F + 10M	10F + 10M
Solvent control (0.001% methanol)		10F + 10M	10F + 10M
50		10F + 10M	10F + 10M
500		10F + 10M	10F + 10M
5,000		10F + 10M	10F + 10M
Total sampled per test	16	100	100

[a] Measurements made on each individual were: wet weight, vitellogenin, spiggin, gonado-somatic index, and gonad histology. F, female; M, male. All concentrations are ng/L.

Kingdom or other European waters, and second, none has a specific, quantitative, repeatable, and reproducible endpoint for androgens or anti-androgens. Presently, the only indication of possible androgenic effects in the 21 d screen are obtained from evaluating reduced VTG levels in females (all three species) and the induction of secondary sexual characteristics in medaka and fathead minnow. Although the female fathead minnow develops nuptial tubercles in response to exogenous androgens [7], the sensitivity of this endpoint and the ability of different laboratories to measure it quantitatively, repeatably, and reproducibly has been brought into question [5]. The only other relevant and reproducible androgen endpoint that has potential among the core species is the development of anal fin papillary processes in female Japanese medaka [5,8].

Research under the United Kingdom Endocrine Disruption in the Marine Environment developed an androgen-specific biomarker using the three-spined stickleback, *Gasterosteus aculeatus* [9]. This biomarker is spiggin, a proteinaceous glue produced by the kidneys of males in response to endogenous androgens and used for constructing nests during the breeding season. Spiggin is also produced by females exposed to exogenous androgens since they also have the androgen receptor and associated biosynthetic pathways (i.e., analogous to male fish producing VTG in response to exogenous estrogens). This work was advanced further by the development of a stickleback VTG assay, providing a specific and sensitive biomarker of estrogenic exposure in male stickleback [10]. As a result, the stickleback has been recognized as a possible additional/alternative test species that could be included in the OECD endocrine disrupter fish screening assay protocol, which is more ecologically relevant for European waters and, more importantly, now has fully quantitative endpoints for both estrogens and anti-estrogens (VTG) and androgens and anti-androgens (spiggin). The ability to measure both endpoints potentially allows simultaneous testing for (anti)-estrogenic and (anti)-androgenic properties of compounds, thereby reducing the number of fish required in experiments (and thus presenting both practical and welfare benefits). A proposal to consider inclusion of the stickleback in the endocrine disrupter screen was presented to the VMG-Eco in 2003, which agreed a small intercalibration exercise should be carried out to evaluate the feasibility of the OECD fish screening assay, adapted for stickleback, and generate data on the reproducibility of the stickleback spiggin and VTG assays, thus bringing information on the stickleback into line with that on the core species. The present paper reports the outcomes of the intercalibration exercise.

MATERIALS AND METHODS

Participating laboratories

Three laboratories participated in the intercalibration exercise: the Centre for Environment, Fisheries and Aquaculture Science (Cefas, UK), which took on the role of lead laboratory, the Centre for Ecology and Hydrology (CEH, UK), and the University of Bergen, Norway. The study took place between March and June 2004.

Stickleback sources

Adult stickleback were obtained separately by each laboratory from wild populations from sources where water quality was relatively pristine, i.e., free from major contaminants (Harrietsham Fish Farm, Sittingbourne, Kent, UK; Moore and Moore Carp, Reading, Berkshire, UK; Drøbak Marine Biological Station, situated on the southern end of the Oslo Fjord, Norway) and acclimated to laboratory experimental conditions for a minimum period of two weeks before exposures began.

Test chemicals and test design

An experimental protocol for stickleback was developed (Table 1), following as far as possible the OECD protocol that was used in the Phase 1A intercalibration with the core species [5].

Two compounds were tested—the natural estrogen E2 and the synthetic androgen 17β-trenbolone. All three laboratories used the same source of test compound: E2 was obtained from Sigma (Gillingham, Dorset, UK), and trenbolone was obtained from Steraloids (Newport, RI, USA). Stock solutions of both were prepared with the use of a carrier solvent (methanol) and dosed under continuous flow conditions (chemical stock flow 0.1 ml/min and dilution water flow 100 ml/min). The following nominal concentrations were used: E2 at 10, 32, and 100 ng/L; and trenbolone at 50, 500, and 5,000 ng/L. A solvent control treatment and dilution water only (negative control) treatment were included in the experimental design. Final concentrations of the solvent in the test and solvent control tanks

were 0.0001% (E2) and 0.001% (trenbolone). Volume exchange of test solutions was 4.8 times per day.

Fish of approximately 1.5 g wet weight ($\pm$ 20%) were used in the exposures and were sexed and separated a priori. Each treatment consisted of two glass aquaria ($\sim$30 L capacity) with 20 males in one tank and 20 females in the other. Mortality was recorded daily and dead fish removed from the tanks. Fish were fed ad libitum with frozen blood-worm (*Chironomus* sp.) once daily. Water temperature was 15°C ($\pm$ 2°C) with a light: dark 12:12 h photoperiod regime. The pH range was 6.5 to 8.5 ($\pm$ 0.5 pH units within a test). No aeration was supplied but water flow coupled with a low stocking density was sufficient to maintain an oxygen concentration >60% air saturation value. Each laboratory recorded physicochemical conditions in each of the test tanks at least once weekly.

Fish were randomly sampled from each tank on day 0 (eight males and eight females from the stock population), day 14 (10 males and 10 females per treatment), and day 21 (remaining 10 males and 10 females per treatment). Fish were anaesthetized with tricaine methane sulphonate (Sigma-Aldrich; 100 mg/L in dilution water buffered with 200 mg NaHCO$_3$/L), blotted dry, and weighed. Blood was collected into a heparinized microcapillary tube after severing the caudal peduncle, transferred into a preweighed 0.5 ml capped centrifuge tube and reweighed (to determine amount of sample obtained from each fish). A 200 μl aliquot of assay buffer (composition determined by Katsiadaki et al. [9]) was added to each sample. The sample was mixed by gentle vortexing and immediately centrifuged at 15,115 *g* for 15 min at 4°C. The clear supernatant (plasma) was pipetted off, placed in a labelled capped centrifuge tube and frozen at -80°C. Following blood sampling, each fish was humanely killed by decapitation and sexed, and the gonads excised and weighed. Kidney tissue was removed, transferred to a capped tube, and frozen at -80°C. Following fixation for circa 48 h in phosphate-buffered 4% neutral formalin, gonads were transferred to 70% industrial methylated spirit in a glass vial.

Gonad histology

Criteria for the histological assessment of gonadal development of male and female sticklebacks were developed by the lead laboratory as part of the present study (Appendix). Using OECD criteria for gonad development in the fathead minnow as a template [11], it was possible to develop and validate similar qualitative criteria in the stickleback. Representative images of each gonadal stage were also taken to increase homogeneity of results amongst participating laboratories. Once developed, these criteria were distributed to the other laboratories to conduct their own assessments.

Each fixed gonad was processed to paraffin wax using standard protocols. Each wax block was sectioned on a microtome (3 μm). To ensure that the qualitative assessment of development was representative, three sections were taken at each of three different steps along the longitudinal axis of the gonad, at each periphery and the center, and stained with hematoxylin and eosin. To confirm that the three step sections were representative of the whole gonad tissue, additional serial sections (three for males, five for females) were taken in between each step and all sections assessed for staging (see Appendix).

Biomarker measurement

All laboratories used a common protocol to measure VTG in plasma and spiggin in kidney via enzyme-linked immu-

nosorbent assays (ELISA), based on previously published methods [9,10,12,13]. The ELISA procedure for VTG is identical to that described in detail for spiggin in Katsiadaki et al. [9]. The differences, described briefly below, are related to the reagents used.

The VTG standard was prepared by lyophilization of the purified protein [10]; for each batch of assays a fresh standard was prepared from the freeze-dried material. Standards range from 0.01 μg/ml (the detection limit of the assay) to 50 μg/ml. Well-plates were coated with stickleback VTG at 100 ng/well. The primary antibody was added at 1:20,000 (v/v) giving a final concentration of the antibody in the pre-incubation mixture of 1:60,000 (v/v). The second antibody and tracer were as for the spiggin assay [9]. The VTG assay has been used for over six years at Cefas using stickleback plasma, whole body homogenates [12], and hearts [13] and has shown high specificity and reproducibility. Different levels of sensitivity can be employed depending on what is expected in the sample under analysis. For the present work, since some high responses were expected in the estrogen-treated groups, the broad range assay was used as described above. However a more sensitive version of the assay can be used, with VTG standards ranging from 0.001 to 5 μg/ml.

The lead laboratory distributed the necessary materials and reagents for the ELISAs to the other participants: polyclonal antiserum against stickleback spiggin, polyclonal antiserum against stickleback VTG, spiggin standard, and purified stickleback VTG standard (lyophilized powder). Measured VTG is expressed as micrograms per milliliter plasma and spiggin as units (U) of spiggin per gram body weight [9]. The limit of detection for the VTG ELISA was 0.01 μg/ml. Taking the dilution factor into account, which varied between individual samples (200 μl of assay buffer was added to each plasma sample before analysis, regardless of the volume), the lowest concentrations measured were approximately 0.1 μg/ml. (This approach was improved following the present study, and now a constant dilution factor is applied, taking into account the volume of plasma collected from individual fish.) The limit of detection for the spiggin assay was 0.8 U/ml of kidney homogenate. However, when this is expressed in spiggin units per gram of body weight, values up to 80 U/g are regarded as below the limit of detection, due to multiplications needed to normalize the samples with the standard dilutions [9].

Analytical chemistry

A sample of water from each tank was removed weekly (on days 0, 7, 14, and 21 unless otherwise stated) for E2 or trenbolone analysis. Water samples were either extracted immediately (using dichloromethane for both compounds) or stored frozen until extraction could take place. Extracts were stored frozen until analysis. Each laboratory was responsible for conducting its own measurements of E2 concentrations in the water via radioimmunoassay, using a standard protocol [14]. At the University of Bergen, although samples were taken from the tanks during the E2 exposure, no chemical analysis was possible due to unavailability of suitable equipment for extraction and subsequent analysis. Since trenbolone was not routinely analyzed in any of the participating laboratories, the chemical analytical method for this test compound required some method development, which was undertaken by the lead laboratory as part of the present study. Subsequently, analysis of trenbolone in water extracts from all three laboratories was performed by Cefas, using liquid chromatography interfaced

Table 2. Measured concentrations of 17β-estradiol and trenbolone from each treatment[a]

	Control	Solvent control	10	32	100
17β-estradiol					
Cefas	—	0.61 (0.49)	8.33 (6.84)	19.58 (3.47)	67.68 (18.58)
CEH	0.38 (0.11)	0.35 (0.18)	4.16 (3.35)	11.5 (6.03)	49.13 (24.56)
	Control	Solvent control	50	500	5,000
Trenbolone					
Cefas	—	—	15 (4)	213 (144)	3,550 (2,363)
CEH	—	6.87 (4.07)	32 (28)	226 (136)	4,646 (3,053)
Bergen	—	—	39 (14)	215 (33)	3,041 (502)

[a] Values for 17β-estradiol are the mean of four sampling occasions (days 0, 7, 14, and 21) during the exposure period. Values for trenbolone are the mean of three (Cefas, The Centre for Environment, Fisheries and Aquaculture Science, Essex, UK: day 7, 14, and 21; the University of Bergen [Bergen, Norway]: day 10, 14, and 21) or four (CEH, Centre for Ecology and Hydrology, Lancaster, UK: day 0, 7, 14, and 21) sampling occasions during the exposure period. Concentrations are expressed as ng/L (standard deviations in parentheses).

with electrospray ionization and a single quadrupole mass spectrometer. Selected ion monitoring was performed on the protonated molecule and selected ion monitoring data were collected in a retention time scheduled event. Quantification was achieved internally using five levels of calibration, and 17-α methyl testosterone as the internal standard.

Data analysis

Data for VTG, spiggin, and GSI in males and females for the two exposures (E2 and trenbolone) and for each of the two sampling occasions (14 and 21 d) was submitted to the lead laboratory for collation and analysis. Where the incorrect sex of fish was discovered in a tank on dissection, the data associated with that individual were pooled with the appropriate sex/treatment and statistical analyses were performed on these group data. In all cases these data were not outliers in their transferred group and did not significantly alter the group mean. The a priori sexing of the stickleback was generally successful. On the majority of sampling occasions all fish were correctly sexed; on a few occasions was one fish of the incorrect sex found. The maximum number of incorrectly sexed fish per 10 dissected was four, on three sampling occasions. The VTG and spiggin raw data were assessed for variance equality by a plot of residuals against fitted values from a

saturated analysis of variance model using laboratory, day, and treatment as factors. Normality was assessed by a quantile-quantile plot of the residuals, i.e., a plot of the empirical quantiles against the theoretical quantiles from a normal distribution. Since raw data demonstrated nonequal variances and nonnormality, VTG and spiggin values were firstly log transformed before analysis using a general linear model procedure (analogous to an analysis of variance, but for unequal sample sizes per treatment) followed by a two-tailed Dunnet's test, in order to assess the ability of each laboratory to detect increases or decreases in the endpoints measured. All treatments (including the negative control) were compared with the solvent control. In only one of the 24 analyses (two exposures·three laboratories·two endpoints·two sexes) did the negative control and the solvent control differ significantly (trenbolone exposure; CEH day 14; spiggin in males endpoint). The statistical software package used was SAS® (Ver. 8, SAS Institute, Cary, NC, USA).

Gonado-somatic index data were not analyzed statistically since no concentration-related effects were evident.

RESULTS

Water chemistry

Measured concentrations of E2, as a proportion of nominals, ranged between 0.60 to 0.68 and 0.31 to 0.47 for Cefas and CEH respectively (Table 2). Measured concentrations of trenbolone, as a proportion of nominals, ranged between 0.28 to 0.86, 0.31 to 1.21, and 0.42 to 0.81 for Cefas, CEH, and the University of Bergen, respectively (Table 2). The 500 ng/L female treatment at Cefas produced particularly low recovery rates on days 14 (one-fifth of nominal) and 21 (one-tenth of nominal); the reasons for this are not clear, but were not due to pump failure or dosing volume inaccuracy.

Although measured levels of both E2 and trenbolone were reasonably (with a few exceptions) stable, they were generally below the nominal values, indicating that the true no-observed-effect concentrations (NOEC) and lowest-observed-effect concentrations (LOEC) were lower than reported below.

E2 exposure—VTG endpoint

Mean baseline concentrations of VTG in males (i.e., concentrations typically found in the day 0 control and in the experimental controls) were low, and generally <4 μg/ml plasma (Fig. 1). At day 14, one of the three laboratories detected a significant increase in VTG at 32 ng/L (nominal), but this
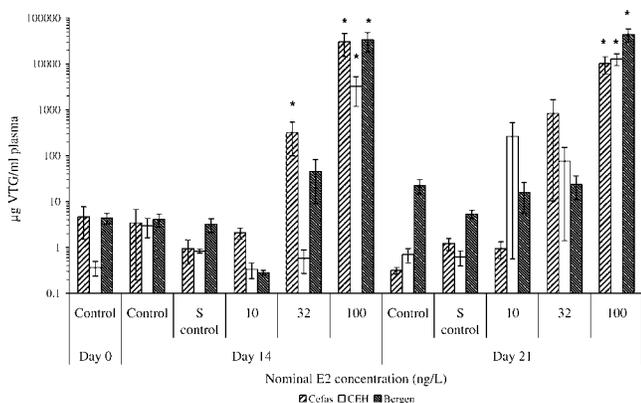

Fig. 1. Vitellogenin (VTG) induction in male stickleback exposed to E2 (17β-estradiol) for 14 and 21 d. Values plotted as means ± standard error. * denotes significantly different (*p* < 0.0001) from the corresponding solvent control (S control). Cefas, The Centre for Environment, Fisheries and Aquaculture Science, Essex, United Kingdom; CEH, Centre for Ecology and Hydrology, Lancaster, United Kingdom; Bergen, University of Bergen, Bergen, Norway.

was not the case after 21 d exposure, although VTG did increase in a concentration-dependent manner. For both days 14 and 21, all three laboratories demonstrated a significant increase in plasma VTG in fish exposed to 100 ng/L (nominal). Not taking into account the outlying result, the NOEC and LOEC were therefore 32 and 100 ng/L respectively at 14 and 21 d for all laboratories. It should be noted that at day 21, CEH detected one male with high VTG at both 10 and 32 ng/L (2,098 and 598 μg/ml, respectively) and Cefas detected one male with very high VTG (6,617 μg/ml) at 32 ng/L, but this was not sufficient to influence the statistical outcome, and the means were not statistically different from the respective solvent controls. The mean control concentration at the University of Bergen on day 21 was slightly elevated (~20 μg/ml); this was due to three individual males with VTG concentrations between 30 and 50 μg/ml. This was most likely due to contamination during sampling rather than contamination in the ELISA or incorrect sexing, since a repeat of the VTG assay on these samples gave the same results and corresponding spiggin values for these individuals were >15,000 U/g body weight.

Baseline levels of VTG in females were high across all three laboratories, with concentrations generally over 10,000 μg/ml (data not shown), indicating that the females were reproductively mature and entering the breeding season. Exposure to E2 resulted in some concentration-dependent increases in VTG, as would be expected, but taken together the data did not permit reliable NOEC and LOEC values to be identified.

E2 exposure—spiggin endpoint

Baseline levels of spiggin in males were high, in the region of 100,000 U/g body weight (data not shown), indicating that the male fish were reproductively mature and were entering the breeding season. There was no significant concentration-related inhibition of spiggin production by E2 at any of the laboratories. Baseline levels of spiggin in females were, as expected, low and mean values were generally less than 100 U/g body weight (data not shown). There was no concentration-related effect of E2 on spiggin production in females at any laboratory.

Trenbolone exposure—spiggin endpoint

The female negative control treatment at Cefas was affected by a fungal infection between days 14 and 21, causing high mortality and thus day 21 data was excluded from any analyses. Baseline levels of spiggin in females at two of the laboratories were below 100 U/g body weight (Fig. 2). Baseline levels at one laboratory (CEH) were approximately 10 fold higher. All three laboratories were able to detect a statistically significant increase in spiggin after 14 and 21 d exposure to trenbolone at a (nominal) concentration of 5,000 ng/L, thus for both exposure times, the LOEC was 5,000 ng/L and the NOEC was 500 ng/L. At Cefas and the University of Bergen, exposure to 5,000 ng/L resulted in an increase in spiggin concentrations of approximately three orders of magnitude. Because CEH baseline levels were higher, the observed increase at the highest concentration was in the region of two orders of magnitude. The reasons for this difference in absolute values are not clear, and were not apparent in the results for males. Although initially postulated to be due to the later timing of this exposure (late June vs May for Cefas and the University of Bergen) when, theoretically, males would be at a more
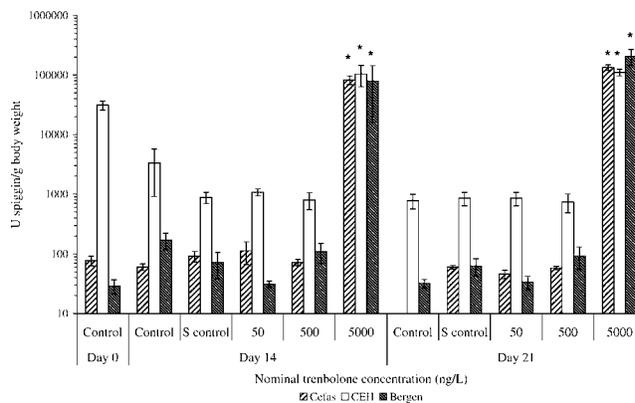

Fig. 2. Spiggin induction in female stickleback exposed to the synthetic androgen trenbolone, for 14 and 21 d. Values plotted as means ± standard error. * denotes significantly different ($p < 0.0001$) from the corresponding solvent control (S control). Cefas day 21 negative control data excluded due to fungal infection. Cefas, The Centre for Environment, Fisheries and Aquaculture Science, Essex, United Kingdom; CEH, Centre for Ecology and Hydrology, Lancaster, United Kingdom; Bergen, University of Bergen, Bergen, Norway; U, units of spiggin.

advanced stage in breeding condition, with subsequent higher levels of secreted androgens in the water, this was not supported by either spiggin concentrations in day 0 males, GSI, or gonad score. Other factors such as stocking conditions prior to the test (e.g., greater stocking density and/or lower flow rates) may have had a bearing on circulating androgen levels.

Baseline levels of spiggin in males ranged from approximately 500 to 100,000 U/g body weight across all three laboratories (data not shown). There was no concentration-dependent effect of trenbolone at the University of Bergen or Cefas—all treatments had high levels of spiggin and the additional effect of the exogenous androgen was minimal. Fish from the CEH exposure showed a concentration-dependent increase at both day 14 and day 21 (data not shown). The increase was significant at 5,000 ng/L on day 14 and 500 and 5,000 ng/L on day 21. Taken together, the data did not permit the identification of reliable LOEC or NOEC values.

Trenbolone exposure—VTG endpoint

Baseline VTG levels in males were low across all laboratories and ranged from approximately 0.15 to 20 μg/ml (data not shown). No concentration-dependent effect of trenbolone on VTG in males was observed.

Baseline levels of female VTG at Cefas were high, >10,000 μg/ml plasma (Fig. 3). After 14 d, mean VTG levels in females exposed to trenbolone at 5,000 ng/L were lower than those in fish exposed to 50 and 500 ng/L, but this trend was not statistically significant. After 21 d a significant inhibition of VTG in 50 and 500 ng/L trenbolone-exposed fish, although not in the 5,000 ng/L exposed group, was evident at Cefas. The results from CEH show variable and low concentrations of VTG in females exposed to trenbolone across all treatments (Fig. 3) with no evidence for a concentration-dependent effect (although at day 21 the concentrations of VTG do decrease in a concentration-dependent manner, the [solvent] control values were lower still). The University of Bergen demonstrated an initial (nonsignificant) increase in VTG at day 14 before a reduction at 5,000 ng/L (Fig. 3). At day 21, there was a concentration-dependent decrease in VTG compared to the control, which was significant at 5,000 ng/L. Taken together, the
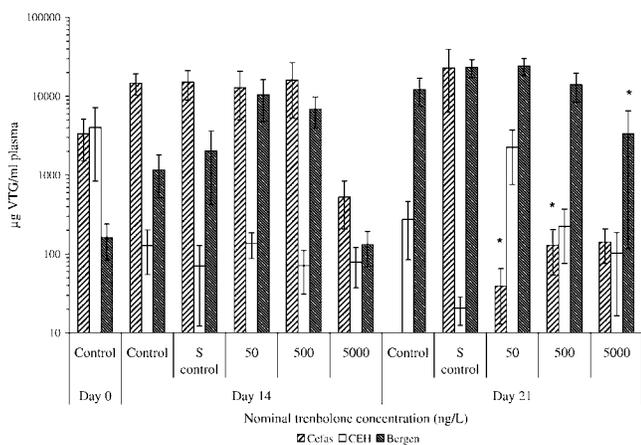
Fig. 3. Vitellogenin (VTG) concentrations in female stickleback exposed to the synthetic androgen trenbolone for 14 and 21 d. Values plotted as means ± standard error. * denotes significantly different ($p < 0.05$) from the corresponding solvent control (S control). Cefas day 21 negative control data excluded due to fungal infection. Cefas, The Centre for Environment, Fisheries and Aquaculture Science, Essex, UK; CEH, Centre for Ecology and Hydrology, Lancaster, United Kingdom; Bergen, University of Bergen, Bergen, Norway.

high variability within the data do not support the identification of NOEC and LOEC values.

Gonado-somatic index

Mean GSI values were calculated for males and females from each treatment in each exposure. There was no consistent pattern in GSIs in either exposure (data not shown), indicating that GSI did not respond to either the E2 or trenbolone. Variation between individual fish was high.

Histopathology

Due to financial and time constraints, the University of Bergen was unable to perform histological analyses of the gonads. Cefas conducted a full assessment and CEH assessed male and female gonad samples from the highest concentration and solvent control of each exposure on day 21 only.

Females

Gonadal stage of development (Appendix) for Cefas females ranged between three and four across all treatments in both exposures. At CEH, all females, including the solvent control treatment, from the E2 exposure were classed as having stage four gonads (i.e., contained many mature oocytes); in the trenbolone exposure all solvent control females were staged as zero or one whereas several trenbolone treated females scored two or three. Although this is most probably a random occurrence, without full histological assessment of other treatments, a treatment-related effect cannot be construed. No treatment-related effects, either pathological or in terms of gonadal maturation, were observed, although at Cefas it was noted in the E2 exposure that across all treatments (i.e., including the control), at day 14 and 21 the number of gonads classed as stage two and one respectively, increased. In the trenbolone exposure, frequent incidences of ovarian atresia (i.e., breakdown of mature oocytes) were observed, particularly after 21 d exposure. However, this was not correlated with the treatments.

Males

For both Cefas and CEH there was no significant change in the testes score for either exposure. All Cefas males had a gonadal development score of three on day 0 for both exposures. In the E2 exposure, 16 and 14% of males exposed to 100 ng/L E2 after 14 and 21 d, respectively, had a lower (stage two) gonadal development stage. A single example of ovotestis (intersex) was observed in a male stickleback exposed to 32 ng/L E2 at Cefas. This was diagnosed as a stage one lesion using the ovotestis severity index devised by Bateman et al. [15]. However, because it was an isolated incidence, it is unlikely that it was the result of exposure to the estrogen. At CEH males had a score of two or three in the solvent control and highest treatment of both exposures.

DISCUSSION

The results of the present study demonstrate that the draft OECD short-term fish screening assay protocol for endocrine disrupters can be successfully applied to an additional species, the three-spined stickleback. The biomarker endpoints of VTG in males and spiggin in females are both relevant and reproducible, changes in the concentration of each reliably reflecting exposure to the potent estrogen E2 and the potent androgen trenbolone respectively.

The VTG endpoint is identical to that used for the core fish species in the OECD protocol for detection of estrogenicity in test chemicals. All three participating laboratories detected statistically significant increases in plasma VTG concentrations in males after 14 and 21 d exposure to 100 ng/L E2 (nominal). This compares very well with similar studies using the core species; 21 d LOECs for E2 have been reported as 100 ng/L for both fathead minnow [1] and zebrafish [16]. The 21 d LOEC for medaka has been reported as 56 ng/L [17]. Seki et al. [8] recently reported lower values for these three species in their study as part of the Phase 1A intercalibration—29, 86, and 9 ng/L for fathead minnow, zebrafish, and medaka respectively (based on mean measured concentrations). These results were mirrored to some extent by other participating laboratories; for fathead minnow and medaka, all (four) laboratories were able to detect significant increases in male VTG after 14 and 21 d at 32 ng/L (nominal); results for zebrafish were less conclusive, with only two out of five labs able to detect increases after 14 and 21 d at 32 ng/L (nominal), and four out of five at 100 ng/L (nominal) [5].

The novel endpoint of spiggin induction in females for detection of androgenicity is unambiguous, quantitative, and unique to the stickleback—it cannot be measured in the existing OECD core species. In the present study, all three laboratories detected statistically significant increases in spiggin in females after 14 and 21 d exposure to 5,000 ng/L (nominal) trenbolone. Because of the unique nature of this endpoint, there are no studies with other species with which to directly compare this threshold value. The main endpoint for androgenicity in the core species is inhibition of VTG in females (an effect which is not androgen-specific) and also, where applicable, observations of masculinization of secondary sexual characteristics in females. In the same study as reported above, Seki et al. [8] observed that the LOEC for VTG reduction in females exposed to trenbolone was 40 ng/L for medaka, 4,060 ng/L for fathead minnow, and 351 ng/L for zebrafish. Similar results were obtained by other laboratories in the intercalibration for fathead minnow (all laboratories detecting a VTG decrease at nominal 5,000 ng/L after 14 and 21 d) and zebrafish (four of

five laboratories detecting a VTG decrease at nominal 500 ng/L after 14 and 21 d). For medaka, a significant VTG decrease was detected by all laboratories only at 500 ng/L. The secondary sexual characteristic of anal fin papillary processes in Japanese medaka appears to be sensitive to trenbolone [5,8], with all laboratories detecting an increase in numbers on females after 14 and 21 d exposure to 500 ng/L. However, this endpoint may allow scope for ambiguous quantification, or subjectivity, particularly in laboratories with less experience with this species.

The effect of trenbolone on VTG in females in the present study was also examined, to compare responses with the core species. The results were less conclusive than those for spiggin production in females and for the VTG results obtained with core species in Phase 1A. One laboratory in the present study measured low baseline concentrations of VTG in females and therefore it was not possible to determine any inhibitory effect of the androgen. The other two laboratories measured relatively high baseline concentrations of VTG and there was some evidence for a decrease in VTG with increasing trenbolone concentration after 14 d, but results were only significant after a 21 d exposure. One laboratory detected a decrease at the lowest tested concentration whereas the other laboratory only detected a decrease at the highest concentration. Based on these results, together with the fact that there exists a superior, androgen-specific endpoint in spiggin production, it is concluded that VTG depression in females cannot be considered to be a diagnostic or reliable endpoint for the detection of androgens in a screening assay with stickleback.

On first comparison of the core species' results with those obtained in this exercise, it would seem that the stickleback is slightly less sensitive to both test chemicals. It should be borne in mind, however, that the chemistry results from Phase 1A indicate that, for all participating laboratories that submitted data, measured concentrations were very close to the nominal values (on average 80% or greater) [5]; in the present study, the measured concentrations were generally lower (and in some cases, substantially lower) than nominals (e.g., Cefas trenbolone after 14 and 21 d). In the present study, measured E2 concentrations, as a mean across labs, were 70% of nominals, and trenbolone averaged 58% of nominal. If these actual values were employed, the LOEC for E2 would be in the region of 50 to 70 ng/L rather than 100 ng/L, thus minimizing the apparent difference in sensitivity between the stickleback and the core species. Indeed, other studies performed at Cefas with stickleback (I. Katsiadaki and Y. Allen, personal communication, unpublished results) have demonstrated that VTG can be induced at nominal concentrations of 32 ng/L E2 after 21 d. It is observed that, on average, actual E2 concentrations were greater at Cefas than CEH and this is mirrored by the mean VTG concentrations. This supports the value of measuring actual concentrations in these types of test. In the case of trenbolone, the NOEC for spiggin induction in females may also not be as high as it appears, since the actual concentrations were lower than 500 ng/L. The LOEC may be reduced to approximately 3,000 ng/L if actual concentrations were to be used, although this reduction is not sufficient to bring the LOEC significantly closer to that observed for the core species. It is worthy to note that, despite Cefas having experienced a problem with trenbolone measured concentrations in the latter part of the exposure, the spiggin concentrations in females remained unaffected. While the 500 and 5,000 ng/L treatments displayed seriously depleted trenbolone concentrations be-

tween days 14 and 21, no corresponding decrease in spiggin was observed, concentrations in females remaining significantly increased. It is not known at exactly which point between the two sampling periods that problems began, or how intermittent or sustained they were, but it is clear that during the 7 d window the spiggin concentration was maintained despite a lowered stimulus, indicating that, once induced, high levels can be maintained for some time and the half-life is greater than 7 d. Unpublished studies have shown that at least one month is needed for a fully hypertrophied kidney to return to a nil spiggin signal (I. Katsiadaki, personal communication, unpublished data). This value is based on observations of breeding males sampled after the photoperiodic stimulus was removed and from studies on anti-androgens where total absence of spiggin was detected after 21 d of exposure in all males. It may be the case that the measured concentration of approximately 1,000 ng/L trenbolone was sufficient to keep the spiggin levels topped up.

Another potential reason for the differences in sensitivity may be the differences in test temperatures. The OECD core fish are all warm-water species and recommended test temperatures are 24°C and above. The stickleback, on the other hand, is a cold-water species and the test is run at 15°C. It is plausible therefore, that differences in metabolic rates lead to a slower response to the estrogen/androgen in stickleback. Given a longer exposure duration, the stickleback LOECs may be reduced such that they are comparable to the core species. Certainly, it has been observed from longer-term exposures (I. Katsiadaki, personal communication, unpublished data) that the VTG response does not plateau after 21 d but continues to increase. It would also be interesting to conduct stickleback exposures at higher temperatures, within their tolerance range, and measure the same endpoints.

Gonado-somatic index was shown not to be a relevant or reliable endpoint for the detection of estrogenic or androgenic activity, which is in agreement with the results obtained for the core species. No concentration-dependent effects on GSI were observed, indicating that gonadal development was not affected by either exposure.

As part of the present study, histological criteria for staging stickleback gonads were developed. The two laboratories that were able to conduct such assessments concluded that there were no significant effects as a result of exposure to the estrogen or androgen. Only one intersex fish was found from the entire dataset. In the Cefas E2 exposure, the increased occurrence of female fish with a lower score across all treatments is unlikely to be due to regressed maturation; rather, it could be attributed to the summer photoperiod which caused some females to ovulate (this is possible in the absence of males [18]) and subsequently invest again in stages of early vitellogenesis. However, this was not observed in the trenbolone study, where an increased incidence of atresia across all treatments indicates that females were unable to ovulate. The reasons for this difference between the two exposures are not clear. It was initially hypothesized that a greater number of rogue males in the female tanks in the E2 study stimulated more females to ovulate, but this was not the case. Without a full histological dataset from the other laboratories it is difficult to interpret the data further.

At the time of conducting this first intercalibration, an insufficient level of guidance and standardization of procedures to evaluate, in a harmonized and comparable fashion, core species gonad samples existed, which would have been used

as a base for developing parallel stickleback histological criteria. For this reason, no firm conclusions about the suitability of the histology endpoint in the screen could be made for either the core species or the stickleback after this first round of validation. To address this, a consensus guidance document was subsequently produced by a group of expert pathologists for use in Phase 1B (as an appendix to the Phase 1B protocol), focussing on four key assessment criteria for each sex; this is currently being applied in a second intercalibration with the stickleback that mimics the Phase 1B intercalibration with the core species. The application of qualitative and quantitative histological measurements for the assessment of gonad tissues of fathead minnow following exposure to EDCs is effective [19]. The approach used by these authors requires a more rigorous methodology than that adopted during the present study, using Bouin's fixative and wax embedding for ovary and glutaraldehyde fixations and resin embedding for testis. Categorization and quantification of cell types from digitized images using image analysis software enabled statistical comparison of exposure groups. The quantitative assessment adopted by Wolf et al. [19] will also be appropriate for use with the other small fish species.

With regards to the lack of evidence for expected antagonistic effects of the steroids tested (i.e., VTG suppression in females exposed to the androgen and spiggin suppression in males exposed to the estrogen), we can surmise that this was perhaps due to the time of year that the exposures took place. Stickleback have a very strong seasonal cycle [18] and the conditions of the test (and acclimation period) would have stimulated the production of natural steroids which in turn induce high baseline concentrations of VTG in females and spiggin in males. It is probable that these high levels of natural steroids were responsible for overriding any down-regulating effect of the added compounds. Since the stickleback endpoints for detection of estrogens and androgens involve receptor-mediated inductions of proteins that are not normally present, the endogenous levels of VTG in females and spiggin in males could be regarded as inconsequential and the assay as suitable for year-round testing. However, this would only apply to a screen that was intended solely for the detection of estrogens and androgens; the OECD fish screen for EDCs has wider applicability and must also be able to detect estrogen and androgen antagonists as well as aromatase inhibitors. Endogenous levels of proteins may therefore be paramount to the success of detecting these classes of compound.

It has recently been demonstrated in a second phase of validation with core species that the androgenic endpoints do not appear to be antagonized by the model anti-androgen flutamide [6] and thus the screen currently lacks the ability to identify anti-androgens. This particular mode of action has become increasingly relevant for the United Kingdom, with a recent survey of sewage effluents indicating that many have a very high level of anti-androgenic activity, as evidenced by in vitro tests, suggesting that anti-androgenic chemicals from a variety of sources are entering the aquatic environment [20]. The stickleback thus has an additional advantage by possession of an endpoint for anti-androgens. Data have been published showing that induction of spiggin by exogenous androgens in female stickleback can be successfully antagonized by flutamide [10] as well as several pesticides [21]. These studies simultaneously exposed fish to an androgen and anti-androgen and focussed on the spiggin endpoint in females. Since a co-exposure regime would not be acceptable within the OECD

EDC screen guideline an alternative test design needed to be devised which would allow the effects of anti-androgens to be detected. In addition, the nonspawning assay used in validation Phase 1A was modified, at the recommendation of the VMG-Eco, to become a spawning assay in Phase 1B. The spawning endpoint is not used as a core diagnostic endpoint to identify EDCs, rather it is a quality assurance measure introduced to ensure that the adults used in the test are in full breeding condition.

To accommodate these two requirements in the aforementioned second intercalibration with stickleback, a new design was developed by Katsiadaki [22]. Since it has been demonstrated that mature males are much less sensitive to flutamide than co-exposed females due to high levels of natural androgen [21], this assay uses pre-acclimated gravid females and males that have not yet entered the breeding phase. The tanks are set up in such a way as to allow natural behaviors such as nest building and spawning. The effect of anti-androgens can be detected at both the biomarker level and the behavioral level; the photoperiod and water temperature in the test are such that under control conditions the levels of spiggin in males increase during the three-week exposure and they successfully progress through the nest building, courtship behavior, and fertilization phases of the reproductive cycle. On exposure to an anti-androgen, natural spiggin induction is inhibited and there is little or no evidence of any of these phases. These behavioral endpoints thus offer the additional benefit, in the screening tier, of allowing potential effects at the population level to be ascertained, which spiggin concentrations per se cannot do. This design has been used successfully to test two anti-androgens (flutamide and the pesticide fenitrothion). Both spiggin and reproductive behavior were significantly altered (I. Katsiadaki, personal communication, unpublished data) at flutamide-equivalent values that have been commonly reported in United Kingdom sewage treatment works effluents [20]. A thorough knowledge and experience of stickleback biology, together with careful planning should ensure that testing can be conducted throughout most of the year.

The results of this first intercalibration were presented to the OECD VMG-Eco. The group approved the findings and requested that the United Kingdom organize and lead on a second intercalibration to mimic Phase 1B, which assessed core species responses to the weak estrogen 4-*tert*-pentyl-phenol, the anti-androgen flutamide and the aromatase inhibitor prochloraz. This is currently underway and the findings will be reported in due course. It is hoped that the results will continue to demonstrate that the stickleback shows potential as a test species and facilitate its inclusion in the final technical guideline.

REFERENCES

1. Panter GH, Thompson RS, Sumpter JP. 1998. Adverse reproductive effects in male fathead minnows (*Pimephales promelas*) exposed to environmentally relevant concentrations of the natural estrogens, estradiol and estrone. *Aquat Toxicol* 42:243–253.
2. Ankley GT, Jensen KM, Kahl MD, Korte JJ, Makynen EA. 2001. Description and evaluation of a short-term reproduction test with the fathead minnow (*Pimephales promelas*). *Environ Toxicol Chem* 20:1276–1290.
3. Panter GH, Hutchinson TH, Lange R, Lye CM, Sumpter JP, Tyler

CR, Zerulla M. 2000. Development of a juvenile fish screening assay for the detection of endocrine active substances. CEFIC-EMSG Aquatic Research Program Progress Report CEFIC-EMSG/AQ002. Brussels, Belgium.

4. Thorpe KL, Hutchinson TH, Hetheridge MJ, Sumpter JP, Tyler CR. 2000. Development of an in vivo screening assay for estrogenic chemicals using juvenile rainbow trout (*Oncorhynchus mykiss*). *Environ Toxicol Chem* 19:2812–2820.

5. Organisation for Economic Cooperation and Development. 2006. Report of the initial work towards the validation of the 21d fish screening assay for the detection of endocrine active substances (Phase 1A). ENV/JM/MONO(2006)27. Series on Testing and Assessment 60. Paris, France.

6. Organisation for Economic Cooperation and Development. 2006. Report of the validation of the 21d fish screening assay for the detection of endocrine substances (Phase 1B). ENV/JM/MONO(2006)29. Series on Testing and Assessment 61. Paris, France.

7. Panter GH, Hutchinson TH, Hurd KS, Sherren A, Stanley RD, Tyler CR. 2004. Successful detection of (anti-) androgenic and aromatase inhibitors in pre-spawning adult fathead minnows (*Pimephales promelas*) using easily measured endpoints of sexual development. *Aquat Toxicol* 70:11–21.

8. Seki M, Fujishima S, Nozaka T, Maeda M, Kobayashi K. 2006. Comparison of response of 17β-estradiol and 17β-trenbolone among three small fish species. *Environ Toxicol Chem* 25:2742–2752.

9. Katsiadaki I, Scott AP, Hurst MR, Matthiessen P, Mayer I. 2002. Detection of environmental androgens: A novel method based on enzyme-linked immunosorbent assay of spiggin, the stickleback (*Gasterosteus aculeatus*) glue protein. *Environ Toxicol Chem* 21:1946–1954.

10. Katsiadaki I, Scott AP, Mayer I. 2002. The potential of the three-spined stickleback (*Gasterosteus aculeatus* L.) as a combined biomarker for estrogens and androgens in European waters. *Mar Environ Res* 54:725–728.

11. Organisation for Economic Cooperation and Development. 2002. Final report of the technical workshop on fish gonadal histopathology. ENV/JM/TG/EDTA/M(2002)1. Paris, France.

12. Hahlbeck E, Katsiadaki I, Mayer I, Adolfsson-Erici M, James J, Bengtsson B-E. 2004. The juvenile three-spined stickleback (*Gasterosteus aculeatus* L.) as a model organism for endocrine disruption II—Kidney hypertrophy, vitellogenin and spiggin induction. *Aquat Toxicol* 70:311–326.

13. Andersson C, Katsiadaki I, Lundstedt-Enkel K, Örberg J. 2007. Effects of 17α-ethynylestradiol on EROD activity, spiggin and vitellogenin in three-spined stickleback (*Gasterosteus aculeatus*). *Aquat Toxicol* 83:33–42.

14. Scott AP, MacKenzie DS, Stacey NE. 1984. Endocrine changes during natural spawning of the white sucker *Catostomus commersoni*. II. Steroid hormones. *Gen Comp Endocrinol* 56:349–359.

15. Bateman KS, Stentiford GD, Feist SW. 2004. A ranking system for the evaluation of intersex condition in European flounder (*Platichthys flesus*). *Environ Toxicol Chem* 23:2831–2836.

16. Brion F, Tyler CR, Palazzi X, Laillet B, Porcher JM, Garric J, Flammarion P. 2004. Impacts of 17β-estradiol, including environmentally relevant concentrations, on reproduction after exposure during embryo-, larval- and adult-life stages in zebrafish (*Danio rerio*). *Aquat Toxicol* 68:193–217.

17. Kang IJ, Yokota H, Oshima Y, Tsurada Y, Yamaguchi T, Maeda M, Imada N, Tadokoro H, Honjo T. 2002. Effects of 17β-estradiol on the reproduction of Japanese medaka (*Oryzias latipes*). *Chemosphere* 47:71–80.

18. Wooton RJ. 1976. *The Biology of the Sticklebacks*. Academic, London, UK.

19. Wolf JC, Dietrich DR, Friederich U, Caunter J, Brown AR. 2004. Qualitative and quantitative histomorphologic assessment of fathead minnow *Pimephales promelas* gonads as an endpoint for evaluating endocrine-active compounds: A pilot methodology study. *Toxicol Pathol* 32:600–612.

20. Johnson I, Hetheridge M, Tyler CR. 2007. Assessment of the (anti-) estrogenic and (anti-) androgenic activities of final effluents from sewage treatment works. Science Report SC020118/SR. Environment Agency, Bristol, UK.

21. Katsiadaki I, Morris S, Squires C, Hurst MH, James JD, Scott AP. 2006. Use of the three-spined stickleback (*Gasterosteus aculeatus*) as a sensitive in vivo test for detection of environmental anti-androgens. *Environ Health Perspect* 114:115–121.

22. Katsiadaki I. 2007. The use of the stickleback as a sentinel and model species in ecotoxicology. In Östlund-Nilsson S, Mayer I, Huntingford F, eds, *Biology of the Three-Spined Stickleback*. CRC, Boca Raton, FL, USA, pp 319–352.

APPENDIX

Stage	Criteria
Staging of female stickleback gonads	
0–Previtellogenic (stickleback <9 mm in length)	Chromatin nucleolar and perinuclear, previtellogenic oocytes exclusively
1–Early development	A distinct oocyte nucleus is visible, containing numerous nucleoli arranged around the nuclear membrane. >90% follicles previtellogenic; cortical alveoli become prominent
2–Mid-development	At least 50% of observed follicles are early and midvitellogenic; cortical alveoli are pushed to the periphery of the oocyte
3–Late development	Maturing follicles show stages of yolk formation appearing as globules within the cytoplasm. Thecal and follicular layers become very distinct
4–Late development/hydrated	Majority of developing follicles are late vitellogenic; follicles are much larger and the yolk appears as a homogenous mass
5–Post ovulatory	Spent follicles; remnants of theca externa and granulosa
Staging of male stickleback gonads	
0–Undeveloped (stickleback <9 mm in length)	Exclusively immature phases (spermatogonia, primary and secondary spermatocytes) with no spermatozoa
1–Early spermatogenic	Immature phases predominate; spermatids and spermatozoa may also be observed
2–Mid-development	Spermatocytes, spermatids, and spermatozoa are present in roughly equal proportions
3–Late spermiogenic	All stages may be observed; mature sperm predominate
4–Spent	Loose connective tissue with some remnant sperm