

Development, standardization and refinement of procedures for evaluating effects of endocrine active compounds on development and sexual differentiation of *Xenopus laevis*

Ilka Lutz · Werner Kloas · Timothy A. Springer ·
Larry R. Holden · Jeff C. Wolf · Henry O. Krueger ·
Alan J. Hosmer

Received: 30 October 2007 / Revised: 7 February 2008 / Accepted: 11 February 2008 / Published online: 8 March 2008
© The Author(s) 2008

Abstract *Xenopus laevis* has been introduced as a model to study effects of endocrine-active compounds (EAC) on development and sexual differentiation. However, variable and inconsistent data have raised questions about the reliability of the test methods applied. The current study was conducted in two laboratories to develop, refine, and standardize procedures and protocols. Larvae were exposed in flow-through systems to 17 β -estradiol (E2), at concentrations from 0.2 to 6.0 $\mu\text{g E2 L}^{-1}$ in Experiment 1A, and 0.015 to 2.0 $\mu\text{g E2 L}^{-1}$ in Experiment 1B. In both studies survival (92%, 99%) and percentage of animals that

completed metamorphosis (97%, 99%) indicated reproducible biological performance. Furthermore, minor variations in husbandry led to significant differences in snout-to-vent length, weight, and gonad size. In Experiment 1A, almost complete feminization occurred in all E2 treatment groups whereas a concentration response was observed in Experiment 1B resulting in an EC_{50} of 0.12 $\mu\text{g E2 L}^{-1}$. The final verified protocol is suitable for determining effects of EAC on development and sexual differentiation in *X. laevis*.

Keywords *Xenopus laevis* · Flow-through exposure system · Standardized protocol · Development · Sexual differentiation · 17 β -Estradiol

I. Lutz (✉) · W. Kloas
Leibniz-Institute of Freshwater Ecology and Inland Fisheries,
Mueggelseedamm 301, 12587 Berlin, Germany
e-mail: ilka.lutz@igb-berlin.de

W. Kloas
Mathematisch-Naturwissenschaftliche Fakultät I,
Institut für Biologie,
Unter den Linden, 6, 10099 Berlin, Germany

T. A. Springer · H. O. Krueger
Wildlife International Ltd,
Easton, MD 21601, USA

L. R. Holden
Sielken and Associates Consulting, Inc.,
Bryan, TX 77802-4015, USA

J. C. Wolf
Experimental Pathology Laboratories, Inc.,
Sterling, VA 20166, USA

A. J. Hosmer
Syngenta Crop Protection, Inc.,
Greensboro, NC 27409, USA

Abbreviations

ASTM	American Society for Testing Materials
BPA	Bisphenol A
DO	Dissolved oxygen
dpf	Days post fertilization
EAC	Endocrine-active compounds
E2	17 β -Estradiol
ELISA	Enzyme-linked immunosorbent assay
EPA	Environmental Protection Agency
EPL	Experimental Pathology Laboratories
IGB	Leibniz-Institute of Freshwater Ecology and Inland Fisheries
MS 222	Tricaine methanesulfonate
NF	Nieuwkoop and Faber
WLI	Wildlife International Ltd
SAP	Scientific Advisory Panel
YAS	Yeast androgen screen
YES	Yeast estrogen screen

Introduction

During the last ten years, the number of chemicals evaluated for endocrine-active effects has increased. Potential exposure of wildlife populations, including fish, amphibians, birds, and mammals, to environmental estrogens has attracted much attention due to the feasible impact of these substances on survival and population sustainability [1–6]. Laboratory studies using a range of wildlife species have demonstrated that physiological processes including development and reproduction, are sensitive to environmental estrogens [7, 8]. Due to the diversity of wildlife, certain species have been used as representatives of the various taxa that are considered in eco-toxicological hazard assessment. One example of a well established standardized animal model is the frog embryo teratogenesis assay (FETAX) [9] using *Xenopus laevis* embryos to determine impacts of chemicals on early development. This species has been investigated focusing on larval stages for studying the hormonal regulation of metamorphosis, development, sexual differentiation, and the impact of endocrine active compounds (EAC) [10–14]. However, the results in the literature for basic biological parameters such as mortality, growth, development, and sexual differentiation vary considerably among laboratories [15–17] making interpretation difficult. For example, it has been reported that the herbicide atrazine feminizes and demasculinizes amphibians [18–21] but other investigators have not been able to replicate these findings [22, 23].

The published data on potential effects of atrazine on gonadal development in laboratory studies and reproductive effects in wildlife populations were evaluated by a United States Environmental Protection Agency (EPA) Scientific Advisory Panel (SAP) in 2003. The SAP identified a number of factors that adversely affected the quality and comparability of data in the open literature on atrazine. These factors included variability in study design, differences in animal husbandry conditions (e.g., feeding rates), excess biomass in treatment chambers, and the effects of these factors on water quality. Other factors that contributed to difficulty in interpreting the results were variability in the response to the positive control substance and the negative control, as well as differences in the use of gross and microscopic pathology terminology, limited statistical power and lack of compliance with Good Laboratory Practice standards. In consequence, the SAP published a White Paper [24] that highlighted inconsistencies and variability among studies and by so doing provided guidance for the development of a standardized study design to evaluate potential effects of EAC on development and sexual differentiation of *X. laevis*.

The purpose of this study was to develop, standardize, and refine a protocol meeting specifications of the 2003 SAP that could be used to assess estrogenic effects on development

and sexual differentiation in *X. laevis*. A team of scientists representing endocrinology, toxicology, chemistry, pathology, and statistics was involved in the creation of the study design, selection of endpoints, and choice of data analysis tools. The team concluded that it was necessary to conduct two independent studies consecutively in two different locations to develop, refine and validate the study design by assessing effects of E2 on gonadal development of *X. laevis* exposed from NF (Nieuwkoop and Faber) stages 46–48 through stage 66 in a flow-through exposure system. Experiment 1A was conducted at Wildlife International Ltd, MD, USA (WLI) in order to develop and standardize conditions for rearing and handling *X. laevis* larvae, operation of a flow-through system and maintaining water quality and environmental parameters. A concentration range for E2 of 0.2, 1.5 and 6.0 $\mu\text{g L}^{-1}$ was chosen to calculate the effective concentration needed to achieve a 50% effect (EC_{50}) for gonadal feminization and to evaluate abnormal development of gonads. In addition, this preliminary study was used to establish the terminology for classifying gross and histological gonadal abnormalities, to cross-train laboratory personnel, and to develop the appropriate statistical analysis methodology. The procedures developed in Experiment 1A were refined and, implemented in Experiment 1B conducted at the Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Germany (IGB).

Material and methods

Test organism

Tadpoles of *X. laevis* were obtained from the commercial supplier, Xenopus I, (Dexter, MI, USA). These larvae (approx. 1200 for each study) originated from ten pairs of adult breeders for each study. Xenopus I reared the eggs until day post fertilization (dpf) 3 and then shipped them to each laboratory.

Rearing

Upon arrival on dpf 4, tadpoles were acclimated to water and temperature conditions (21 ± 2 °C) over a six-hour time period. During the acclimation period larvae were maintained in two 40-L glass aquaria containing 30 L well water (ca. 600 larvae/tank, Experiment 1A) and three 40-L glass aquaria containing 30 L dilution water (ca. 400 larvae/tank, Experiment 1B) derived from the flow-through system. At the initiation of treatment, twenty-five larvae were randomly assigned to each glass aquarium ($n=200$ larvae per treatment group) so that the biological loading rate was always less than $1\text{g L}^{-1}\text{day}^{-1}$. In Experiment 1A exposures were initiated at dpf 6 NF stage 45–46 and in Experiment

1B at dpf 8 when tadpoles had reached NF stage 47–48 as identified by visible hind limb buds [25]. Exposures continued until completion of metamorphosis or dpf 82, whichever came first. Throughout the acclimation period, water temperature was recorded twice daily whereas dissolved oxygen (DO) concentration and pH were measured daily. Experiment 1B was conducted in compliance with the local animal protection committee (G0048/05).

Feeding

Animals were fed Sera Micron (Lot No: 9698; Sera, Heinsberg, Germany), a commercially available food which was reported by the manufacturer to consist of 50.2% protein, 8.1% fat, 4.2% fibre, and 11.0% ash. A sample of Sera Micron was analysed by Lancaster Laboratories (Lancaster, PA, USA) for potential contaminants (metals and pesticides), and all tested contaminants were below their limits of quantification for analysis at that facility. A mixture of diet and dilution water was tested for estrogenicity and androgenicity using YES (yeast estrogen screen) [26] and YAS (yeast androgen screen) assays [27]. Based on these analyses no estrogenic or androgenic activity was detected.

During acclimation of larvae in Experiment 1A, 1 g Sera Micron was added to each tank daily. This amount of food was found to be excessive. In experiment 1B, after temperature acclimation, larvae were transferred to the acclimation tanks where they promptly attached to the glass aquaria walls. Therefore, no food was added on dpf 4. The following day some of the tadpoles were free swimming and 150 mg Sera Micron was added that evening by rinsing through a sieve. The amount of feed was chosen based on the intake of food by the larvae, and care was taken to ensure that feeding did not adversely affect water quality. From dpf 6 to 8 the free swimming tadpoles were fed 200 mg Sera Micron in each tank twice a day. Throughout the remaining course of both studies, tadpoles were fed a suspension of Sera Micron (77 mg L^{-1}) three times daily. The total quantity of food added per tank each day increased from 300 mg/tank on dpf 8 to 900 mg/tank on dpf 33. The quantity of food added to each tank was adjusted periodically in Experiment 1A and on a daily basis in Experiment 1B to account for mortality and removal of tadpoles as they completed metamorphosis.

Study design

Because an almost completely feminized phenotype was observed in Experiment 1A in all treatment groups, the nominal E2 concentrations in Experiment 1B were adjusted to 0.015, 0.2, and $1.5 \mu\text{g E2 L}^{-1}$, respectively. Mid and high concentrations were selected to correspond to exper-

iment 1A and allow comparison of results from the two laboratories.

Replicates

Each study consisted of at least one negative control group plus three E2-treated groups, each comprising eight replicated glass aquaria/treatment groups. Eight glass aquaria were arranged in two clusters of four tanks each. The clusters of negative control tanks and E2 treated tanks were randomly distributed within the environmental chamber in Experiment 1A and within water baths in Experiment 1B. Each cluster was supplied water via the same water flow meter and mixing chamber. Experiment 1A included 16 control tanks that were initially divided into two 8-tank groups designated as negative control and reference control, respectively. The animals from all treatment groups were handled in a blinded fashion so that the biologists and histologists involved in the experiments did not know the treatment applied to frogs being observed. An additional non-blind reference control group was included to provide reference information on rates of spontaneous incidence of gonadal abnormalities for blind histological assessment of gonads from the E2 treatment groups. Histological analysis was conducted on all frogs of Experiment 1A. At the end of Experiment 1A, it was discovered that slightly elevated temperature ($1 \text{ }^\circ\text{C}$) in one cluster of four negative control tanks had resulted in accelerated growth of the frogs in the tanks. Therefore, measurements of weight, length, gonad image area, and age at metamorphosis from those tanks were dropped from analysis, and the four remaining negative control tanks were combined with the reference control group to provide these measurements.

Flow-through system

Both studies were conducted using continuous flow-through systems operating at a flow rate of at least 50 L dilution water per tank per day. This was equivalent to a water exchange rate of approximately seven tank volumes per day. Test tanks consisted of 9-L glass aquaria ($30 \times 20 \times 14.5 \text{ cm}$) containing 7 L water. Dilution water for Experiment 1A was obtained from a well of approximately 40 m depth located on the WLI site. In Experiment 1B the dilution water used was non-chlorinated municipal water. Dilution water samples from Experiment 1A were analysed for possible metal or pesticides by Lancaster Laboratories and water samples of Experiment 1B by Environmental Chemistry and Pharmaceuticals (RCC Ltd., Itingen, Switzerland).

The flow-through systems in the two laboratories were essentially identical. Prior to use, the water was passed through several filters (5, 1, 0.5, and $0.45 \mu\text{m}$) to remove particulate material, a UV-sterilizer, and a final particle

filter (0.45 μm). After being filtered and temperature conditioned, the dilution water was divided into 16 streams. Each stream passed through a flow-control assembly, consisting of a rotameter flow gauge and a needle valve, to a mixing chamber. The flow from each mixing chamber was then split to supply a cluster of four tanks. Each tank received 140 mL min^{-1} dilution water ($\pm 5\%$).

Primary E2 solutions

The primary E2 stock solutions were continuously delivered to each mixing chamber (0.140 mL min^{-1}), where it was vigorously mixed with the dilution water. Primary E2 stock solutions were metered into the mixing chambers by an eight-channel peristaltic pump (MasterflexL/S, Model 752455, Cole-Parmer, IL, USA). The flow-through system (temperature, water, filter pressure, flow-meter), and peristaltic pumps were calibrated four weeks prior to study initiation, and inspected visually twice daily throughout the study.

E2 solutions

Once every seven days a primary E2 stock solution was prepared in amber bottles by dissolving 6 mg E2 (Lot No: 11121AB, CAS: 50–28–2, purity 99.9%, Sigma Aldrich; WLI, Allentown, PA, USA; IGB: Taufkirchen, Germany) in alkaline solution (0.01 mol L^{-1} NaOH) applying moderate heating and constant mixing of the solution overnight similar to procedures employed by the Duluth Laboratory of the USEPA, Office of Research and Development (personal communication by J. Tietge). In Experiment 1A, the working stock solutions were prepared at nominal concentrations of 200, 1500, and 6000 $\mu\text{g E2 L}^{-1}$, and in Experiment 1B at 15, 200, and 1500 $\mu\text{g E2 L}^{-1}$. The quantity of NaOH in E2 test solutions delivered to each tank was negligible, and therefore no NaOH was added to the negative control.

Analysis of E2 concentration

Sampling

In Experiment 1A, water samples were collected from each tank in all treatment groups to confirm the operation of the diluter two days prior to study initiation. In Experiment 1B, E2 concentrations were monitored twice weekly for a two-week period prior to study initiation. After study initiation, water sampling was carried out twice weekly from alternating tanks in two tanks per cluster until study termination. Throughout the studies water samples were collected routinely two days after the freshly prepared E2

solution was connected to the flow-through system and a second time in the same week. At mid-depth from each test tank 190 mL water was collected using a glass pipette and 10 mL methanol (MeOH) was added. Of this solution 20 mL was decanted into a 20 mL amber scintillation vial for E2-enzyme-linked immunosorbent assay (ELISA) analysis. The remaining volume of 180 mL was stored as backup at 4 °C until E2 analysis was completed. Water samples from Experiment 1B were shipped to WLI for E2 analysis. The stability of E2 was confirmed by shipping samples of E2 ($0.1\text{ }\mu\text{g E2 L}^{-1}$) that were prepared and analysed in parallel with the water samples.

Sample analysis

Rapid analysis was performed using ELISA kits (Ecologena, Abraxis LLC, Warminster, PA, USA) to ensure the absence of the test substance in the negative control tanks and to verify the proper E2 concentration in each E2 tank. The ELISA method was based upon the E2 kit manual provided by the manufacturer. The method detection range of this assay was $0.05\text{--}0.5\text{ }\mu\text{g E2 L}^{-1}$. Parallel analyses of water samples for E2 concentration were performed by direct aqueous injection using an Agilent Series 1100 High-Performance Liquid Chromatograph coupled with a MDS Sciex API 3000 tandem mass spectrometer (LC-MS-MS) (Applied Biosystems, Foster City, CA, USA) and MDS Sciex API heated nebulizer ion source (Applied Biosystems) operated in the multi-reaction mode (MRM). Chromatographic separations were achieved using a Keystone Betasil C-18 column ($50\text{ mm}\times 2\text{ mm}$, 3 μm particle size) (Thermo Hypersil, Madison, WI, USA). The method limit of quantitation (LOQ) for HPLC-MS-MS estradiol analyses was defined as $0.005\text{ }\mu\text{g L}^{-1}$.

Environmental conditions

Temperature

In Experiment 1A, all glass aquaria were situated in a walk-in environmental chamber ($21\pm 2\text{ }^\circ\text{C}$). In Experiment 1B, glass aquaria were placed in temperature controlled water baths ($22\pm 1\text{ }^\circ\text{C}$). In both experiments water temperatures were continuously recorded, and temperatures in all glass aquaria were measured weekly using a liquid-in-glass thermometer.

Light

The target light intensity at tank level was 100–500 lux and was measured weekly using a Sper Scientific (AZ, USA) light meter. A photoperiod of 16 h of light and 8 h of dark was used in Experiment 1A, and was adjusted to a 12 h

light, 12 h dark cycle in Experiment 1B. A 30-min low light (around 30% intensity of full light) transition period was installed in both laboratories.

Water quality

Water quality parameters were assessed according to ASTM (2003) [28] and US EPA guidelines. Water hardness, pH, conductivity, and the concentrations of nitrate, and ammonia were measured weekly in one tank from each cluster throughout the experiments. Hardness and alkalinity were measured by titration based on procedures described in “Standard Methods for the Examination of Water and Wastewater” [29]. Nitrate concentrations were analysed with a Hach DR/700 colorimeter (Hach Company, Colorado, USA) and specific conductance was detected using a YSI conductivity meter (WLI: Yellow Springs Instruments Model 33, IGB: 3200, Yellow Springs, OH, USA). pH and ammonia concentrations were measured with an Orion pH/ISE Meter 720Aplus (Thermo Electron, USA). DO concentrations were measured in alternate tanks per treatment three times a week with an Orion Model 850plus dissolved oxygen meter (Thermo Electron). Commencing at dpf 43 in Experiment 1A, and at dpf 36 in Experiment 1B, gentle external aeration was applied by a glass pipette in each test tank to ensure maintenance of an adequate DO concentration.

Maintaining water quality

The tanks required daily cleaning to minimize microbial growth and to maintain adequate water quality. Bio-film that accumulated on the inner walls and bottom of the tanks was scraped off daily, and detritus was siphoned from the tanks. Daily cleaning proved inadequate to maintain appropriate water quality throughout the entire exposure period. Hence, in Experiment 1A all tanks were replaced with clean tanks weekly and in Experiment 1B twice throughout the experiment when siphoning did not lead to satisfactory results.

In-life assessment of tadpoles

Tadpoles were monitored for changes in general health, swimming behaviour and morphological appearance daily. The numbers of dead or moribund tadpoles and tadpoles completing metamorphosis were recorded daily for each tank.

Snout-to-vent length and weight

Each tadpole that had either completed metamorphosis by NF stage 66, or had failed to achieve metamorphosis by dpf 82, was removed from the test tank and euthanized by

immersion in MS 222 (tricaine methanesulfonate) (WIL: Sigma Aldrich, Allentown, PA, USA; IGB: Sigma–Aldrich, Taufkirchen, Germany) (2 g L^{-1} buffered to pH 7.5 with NaHCO_3). Immediately thereafter, the snout-to-vent length was measured to the nearest millimetre. Each frog was blotted dry and then weighed to the nearest milligram. A unique identifier was assigned to each frog that included: study number, in-life laboratory ID, colour code, replicate, and animal number.

Gross pathology

The pleuroperitoneal cavity was opened to expose the visceral organs. The liver, stomach, and intestine were examined using a dissecting stereo microscope (WIL: Olympus SZ61-TR, Olympus America, Melville, NY, USA; IGB: Olympus, SZX7, Olympus, Hamburg, Germany), and unusual variations in size, colour, and/or abnormal structure were recorded. The gastrointestinal tract was excised to expose the ventral surfaces of the kidney and gonads, which were similarly assessed for gross abnormalities. Tadpoles that did not complete metamorphosis at study termination (Experiment 1A: 25, Experiment 1B: 7) were staged according to NF and, the gonad morphology was assessed (if possible), and the carcasses were archived in fixative. Data derived from such animals were excluded from further analysis.

Assessment of gonad morphology

Using a dissecting stereo microscope, gross evaluations of gonad morphology were performed on all animals that reached NF stage 66. To enhance visualization of the gonads, which presented as thin, pale-tan strips of tissue on the ventromedial margin of the kidney, several drops of Bouin’s solution (WIL: Sigma Aldrich, Allentown, PA, USA; IGB: Sigma–Aldrich, Taufkirchen, Germany) were applied to partially fix these organs. Based on gross observation, the gonads of each animal were identified as testes, ovaries, or malformed. Examples of observed malformations included mixed sex, intersex, pearling and segmental aplasia. Mixed sex was defined as the co-occurrence of both ovarian and testicular tissue in a single gonad. Intersex was assessed as ovarian and testicular tissue in the same individual as separate gonads (left/right). The term pearling was characterized by the presence of multiple, prominent segmental enlargements and/or attenuations along the length of one or both gonads, whereas gonads that presented as nodular islands of testicular or ovarian tissue, with either intervening membranous connections or no connections at all, were classified as segmental aplasia. Gross findings for the gonads of each frog were recorded separately for the left and right gonads.

The gonadal findings for each frog were verified independently by a second biologist. If the two biologists differed in their interpretations, the disparate opinions were discussed and a consensus finding was recorded.

Prior to the initiation of the first E2 experiment a list of standardized terms for changes of potential morphological gonadal observations was developed. Subsequent to the completion of both experiments the applied terminology was refined in order to better characterize and group different types of gonad changes observed and which in turn could be analysed. For example, the subcategories narrow, slightly narrow, truncated, slightly truncated, and margin entire were integrated into the main category hypoplasia. This term describes all gonads which appear smaller than those characterised as “typical/normal”. Since the term “typical/normal” is rather general, further sub-categories allow the biological variability of gonadal morphology to be described. However, during the investigations it was revealed that too few observations in each sub-category ruled out reasonable statistical analyses. Therefore, a refined list of qualitative descriptive terminology for gonadal features was created as presented in Table 1.

Photography

After gross gonad inspection was complete, each gonad was photographed in situ using a digital camera (WIL: Olympus DP12–2, Olympus America, Melville, NY, USA; IGB: Olympus DT5, Hamburg, Germany) attached to the stereo microscope. Each digitalized image included a millimetre measurement scale placed adjacent to the gonads to permit gonad size to be determined from the photograph. Following photography, each frog was placed in a labelled individual container of 30 mL of Bouin’s solution for approximately 48 h. At the end of the fixation period, each carcass was rinsed several times in 70% ethanol and was placed in 30 mL 10% neutral buffered formalin (WIL: VWR, Westchester, PA, USA; IGB: Histofix, Roth, Karlsruhe, Germany). A complete gonadal histological evaluation was performed on each frog of experiment 1A and results are presented in a separate paper (Wolf et al., in preparation). In some cases the growth and histological findings differ and a better understanding of results is gained by considering both types of observation.

Gonad measurement

The biological variability of morphological features of the gonads and the subjective nature of the assessment prompted us to develop a more quantitative metric of gonad size. Thus, individual measurements of gonad image area were derived from photographs of the gonads. Gonad image areas were obtained using image-processing software

(Image Pro Plus, Version 5.1, Media Cybernetics, Silver Spring, MD, USA) that calculated the combined area of the left and right gonad whose outline had been manually traced around the digital image. The program was calibrated by the millimetre measurement scale prior to measurement for the gonad of each tadpole.

Statistical analysis

For both experiments, observations were collected for each animal. However, consistent with the experimental design, the tank is considered the primary experimental unit. When tank differences are present, statistical analysis should accommodate this experimental structure directly using nested “random effects” analyses or indirectly by analysing tank means or tank percentages. Neither approach is completely satisfactory for all endpoints. Nested models have specific distributional requirements and, in the case of incidence data, become unstable when frequencies are at or near zero. Analyses of tank means can also be problematic for these experiments. Estradiol-induced feminization causes the number of males per tank to be smaller for treated groups than for the control group. This, in turn, results in greater variation between estradiol-treated tanks than between control tanks. Most parametric and non-parametric statistical procedures, however, require equal variability between tanks within every group. When tank differences are small or absent, however, the individual data from all tanks within the same treatment group can be pooled and analysed more simply using the frog as the basic experimental unit.

For Experiments 1A and 1B, endpoints were tested for differences between tanks using a test of homogeneity. A chi-square homogeneity test was used for frequency endpoints (males, females, mixed sex) and a Kruskal–Wallis test was used for measurement endpoints (snout-to-vent length, body weight, age at completion of metamorphosis, gonad image area). This homogeneity test was calculated for each experimental group and the *p*-values combined using Tippett’s minimum-*p* method. Tank differences were found for several frequency endpoints in Experiment 1A. In addition, there was evidence that the presence of a humidifier caused a temperature-induced impact on measurement endpoints for two clusters of tanks. In contrast, Experiment 1B did not reveal any statistically significant ($p < 0.05$) tank effects. Consequently, all subsequent statistical analyses for Experiment 1B can be conducted on individual animals with all tanks pooled. Thus, analyses of Experiment 1A are quite complex and varied while Experiment 1B results can be analysed rather simply. In order to summarize the results consistently to allow comparisons of both experiments, only animal-level means and standard

Table 1 Glossary of terms used to describe features observed during gross morphological examination of gonads in *X. laevis* NF stage 66

Feature	Definition
1. Adhesion (other tissue)	Gonads joined to other abdominal tissue
2. Aplasia (agenesis)	Complete lack of gonad development
3. Segmental Aplasia	Gonad is longitudinally discontinuous
3-A. Tissue separation	One or more areas along the length of the gonad is poorly developed, attenuated with an essentially complete lack of gonadal tissue
3-B. Extraneous gonadal tissue	A small disjunct cluster of gonad-like tissue
4. Bifurcation	Division of the gonad oriented longitudinally, along the cranial–caudal axis
4-A. Protuberance(s)	Small projection(s) of the gonad along the margins
4-B. Symmetric	A “Y” -shaped single gonad with relatively equal division
5. Angular deformity	One or more bends in the gonad to an excessive degree
5-A. Gonad folded	The gonad is bent over or doubled up so that one part lies on another
6. Displaced	The gonad, or a section of it, is not located in the typical location
6-A. Displaced laterally	The gonad, or a section of it, is not located in the typical location. Instead it is located further away from the frog’s midline
6-B. Displaced medially	The gonad, or a section of it, is not located in the typical location. Instead it is located closer to or crosses over the frog’s midline
6-C. Displaced cranially	The gonad is not located in the typical location. Instead it is located more towards the anterior
6-D. Displaced caudally	The gonad is not located in the typical location. Instead it is located more towards the posterior
7. Fused	Left and right gonads joined to varying degrees at one or more locations along cranial–caudal axis
8. Hypertrophy	Size of the gonad is larger than typical
8-A. Wide	Gonad appears broad. Lateral width is large relative to the length
8-B. Elongated	Gonad appears long (cranial–caudal). Length is large relative to the length of the kidneys and the torso of the frog
8-C. Thick	Gonad relatively increased in the distance between the dorsal and ventral surfaces
8-D. Enlarged	A generalized increase in the size of the gonad along all axes
9. Segmental hypertrophy	A gonad where one or more areas are excessively large
9-A. Enlargement (mass)	A section of the gonad substantially bigger in width and/or thickness than the remaining tissue
9-B. Pearling	Pronounced multiple segmental enlargement of the gonad
9-C. Partly thick	A section or sections of the gonad is relatively increased in the distance between the dorsal and ventral surfaces
10. Hypoplasia	Size of the gonad is decreased
10-A. Narrow	Gonad appears strap-like. Lateral width is markedly reduced relative to the length
10-B. Slightly narrow	Lateral width is somewhat reduced relative to the length
10-C. Truncated	Gonad appears short (cranial–caudal). Length is small relative to the length of the kidneys and the torso of the frog
10-D. Slightly truncated	Gonad appears somewhat short (rostral–caudal). Length is somewhat reduced relative to the length of the kidneys and the torso of the frog
10-E. Thin	Gonad relatively reduced in distance between the dorsal and ventral surfaces
10-F. Margin entire	All of the ovary or indeterminate gonad has a smooth edge; no scalloping or lobes
11. Segmental hypoplasia	A gonad where one or more areas are excessively reduced, attenuated, or poorly developed but not separated
11-A. Partly narrow	A section or sections of the gonad has a small lateral width relative to the length and the width of the other sections
11-B. Partly thin	A section or sections of the gonad is relatively reduced in distance between the dorsal and ventral surfaces
11-C. Margin slightly sinuate	Gonad has only shallow waved or scalloped edge (typically associated with a narrow or partly narrow ovary)
11-D. Margin partially entire	A section or sections of the gonad has a smooth edge (typically associated with a narrow or partly narrow ovary)
11-E. Pearling	Pronounced multiple segmental attenuation or narrowing of the gonad
12. Intersex	Ovarian and testicular tissue present as separate structures (i.e. not contained in the same gonad) (left/right)
13. Mixed sex	Ovarian and testicular tissue present in the same gonad
14. Translucent	Gonad appears not so dense, light able to pass through diffusely
14-A. Tissue slightly translucent	To a small extent the gonad appears not so dense, a small amount of light able to pass through diffusely
15. Segmental translucence	A section or sections of the gonad appears not so dense, light able to pass through diffusely
16. Melanophores decreased	Apparently fewer pigment-containing cells than typically seen. Applies only to females, mixed sex frogs, and frogs of undetermined sex

Table 1 (continued)

Feature	Definition
16-A. Apigmentation	No melanophores in the internal tissue. Applies only to females, mixed sex frogs, and frogs of undetermined sex
17. Internal melanophores	Pigment-containing cells within the gonadal tissue. Applies only to males, mixed sex frogs, and frogs of undetermined sex
18. Other	Any finding not listed. Describe as needed. For any “other” finding that study personnel consider not incidental, notify the Study Director and other laboratories as soon as possible via email including a description and photograph

deviations are presented. The means and standard deviations given in this report for both experiments were calculated from individual animals using Excel Version 2000. For Experiment 1B these simple statistics are identical with those obtained from the statistical analyses described below. For Experiment 1A these means (and standard deviations) should be considered only approximate. They may differ slightly from means obtained using more sophisticated analyses that incorporate a tank structure.

In Experiment 1B, the estradiol-treated groups were compared with controls in a step-wise manner that preserved power and protected against excessive false positives. An overall test of group differences and a test for trend with dose were both conducted. If either of these two overall tests showed statistical significance at the 5% level (i.e., $p < 0.05$) then comparisons of each dose group to the control were made. For frequency endpoints a chi-square homogeneity test and a Cochran–Armitage test for trend were used for the overall homogeneity and trend tests, respectively. Comparisons of dose groups to control were conducted using Fisher’s exact test. For measurement data essentially all endpoints showed statistically significant deviations from normality ($p < 0.01$, Shapiro–Wilk test). Consequently, a Kruskal–Wallis nonparametric test was used for group homogeneity and a Jonckheere–Terpstra test was used for trend. Comparisons of dose groups to control were conducted using the nonparametric Wilcoxon–Mann–Whitney U-test. All statistical analyses of Experiment 1B were calculated using Release 9.1 of the SAS statistical package (2004).

Results

It is emphasized that Experiment 1A was performed first in order to standardize conditions for rearing *X. laevis* and for operation of the flow-through system. Improvements of these procedures were implemented in Experiment 1B resulting in optimised growth and development conditions for *X. laevis* tadpoles. Those parameters that were successfully achieved during Experiment 1B are summarised in

Table 2 and may be used as guidance for similar experiments. The following sections focus on results obtained in Experiment 1B, presenting results from Experiment 1A, as needed, to allow comparison of the two experiments.

Environmental conditions

In both experiments experimental conditions generally met the criteria listed in Table 2. In Experiment 1B the mean and SD for hardness and alkalinity were 103 ± 5.8 mg and 162 ± 6.8 mg $\text{CaCO}_3 \text{ L}^{-1}$, respectively. Specific conductance was 794 ± 18.0 $\mu\text{S cm}^{-1}$, pH 8.1 ± 0.09 , nitrate 1.6 ± 0.5 mg L^{-1} and ammonia 0.08 ± 0.09 mg L^{-1} . The measurements from negative control tanks were not significantly different from those from the E2 tanks. Technical problems in Experiment 1B prevented accurate determination of DO concentration, however, based on subsequent studies using the same procedures, DO was consistently above 60% saturation.

E2 exposure verification

In each experiment, E2 was not present in any of the negative control tanks. Measured concentrations of E2 stock solutions ranged from 96.6 to 106% in Experiment 1A and between 90.5 and 108% in Experiment 1B. In Experiment 1A, the mean concentration of E2 for three treatment groups began to decline after experiment initiation and fell below 80% of nominal E2 concentration by dpf 29. Between exposure dpf 43 and 50 mean concentration of E2 decreased to minima of 0.4, 43, and 38% of nominal concentrations of E2 in the 0.2, 1.5, and 6.0 $\mu\text{g E2 L}^{-1}$ treatment groups, respectively. After dpf 50, concentrations of E2 began to increase in all treatment groups, and exceeded 60% of nominal concentration on dpf 64 (Fig. 1a). In Experiment 1B, concentrations of E2 declined after study initiation and dropped below 60% of nominal E2 concentration by dpf 23. On dpf 37 concentrations of E2 fell to 47, 46, and 50% in the 0.015, 0.2, and 1.5 $\mu\text{g E2 L}^{-1}$ treatment groups, respectively. The concentration of E2 increased to 60% of nominal concentration on dpf 65 (except the 1.5 $\mu\text{g E2 L}^{-1}$ with 44.4%) (Fig. 1b).

Table 2 Final study design specifications. Brief summary of the most important parameters for exposing *X. laevis* tadpoles in a flow-through exposure system

Parameter	Characteristic
Water-quality parameters	
Supply water: filtered, contaminant free	Filtered and UV sterilized; tested for chemical contaminants
Dissolved oxygen	≥ 60% of saturation
pH	7.9 to 8.3
Ammonia	0.0 to 0.35 mg L ⁻¹
Nitrate	0.04 to 2.29 mg L ⁻¹
Hardness	90 to 130 mg CaCO ₃ L ⁻¹
Alkalinity	133 to 172 mg CaCO ₃ L ⁻¹
Specific conductance	726 to 817 μg S cm ⁻¹
Supplemental aeration	Initiated not later than dpf 36
Animal husbandry	
Animal supply	Xenopus, Dexter, MI, USA
Shipped from Supplier	Day 3 post fertilization (dpf)
Acclimation	dpf 4–7
Treatment period	dpf 8–82
Feed: Sera Micron	Sera Micron (contaminants and estrogenic potential evaluated) Contents: 50% protein, 8% fat; 4% fibre; 11% ash.
Feeding rate—acclimation	dpf 4–5; per 300 larvae: 200 mg; if free swimming, dpf 6–8: 200 mg twice daily (based on the intake of food by the larvae)
Feeding rate—NF stages 46–66	77 mg L ⁻¹ : fed three times daily, 300 mg/tank from dpf 8, increasing to 900 mg/tank at dpf 33, then adjusted during experiment as animals were removed (dead, froglets)
Exposure chamber cleaning	Daily
Exposure chamber replacement	As needed
Water-quality parameters	Hardness, ammonia, nitrate, pH, specific conductivity twice a week; DO three times a week
Temperature	22±1°C (verified twice daily, measured once a week)
Light intensity	100–500 lux (measured once weekly, at level of surface of water)
Light:dark cycle	12:12; 30-min transition (30% of full light)
Experimental design	
Performed in compliance with Good Laboratory Practice Standard	Yes
Colour coded (blind) treatment	Yes
Randomization of tadpoles	Tadpoles distributed in rotation one or two at a time to transport vessels to contain 25, then vessels were randomly assigned to test tanks
Randomization of tank clusters	Completely randomized placement; one cluster consists of four tanks
Negative control	Water (eight replicates, two clusters)
Treatment 17β-estradiol concentrations	0.015, 0.2, 1.5, 6.0 μg E2 L ⁻¹ ; eight replicates; two clusters for each group
Tank volume	9 L (water volume 7 L), tanks covered
Tank flow rate	50 L day ⁻¹
Number of animals	200 larvae per treatment group
Animal load density	Less than 1 g L ⁻¹ water; 25 larvae/tank
HPLC–MS–MS verification	dpf 8 all tanks; dpf 16–81 weekly (alternating tanks)

In-life assessment of tadpoles

Appearance and behaviour

No effects of E2 treatment on tadpoles condition, behaviour, or external appearance was observed in any of the treatment groups at any time throughout the entire course of Experiment 1A or Experiment 1B.

Survival

Excellent survival was achieved in both experiments in controls as well as in E2 treated groups. The total survival was 92% in Experiment 1A and 99% in Experiment B. In Experiment 1A, survival in the combined negative control, 0.2, 1.5, and 6.0 μg E2 L⁻¹ treatment groups was 90.8, 94.5, 95.0, and 88.0%, respectively. In Experiment 1B survival

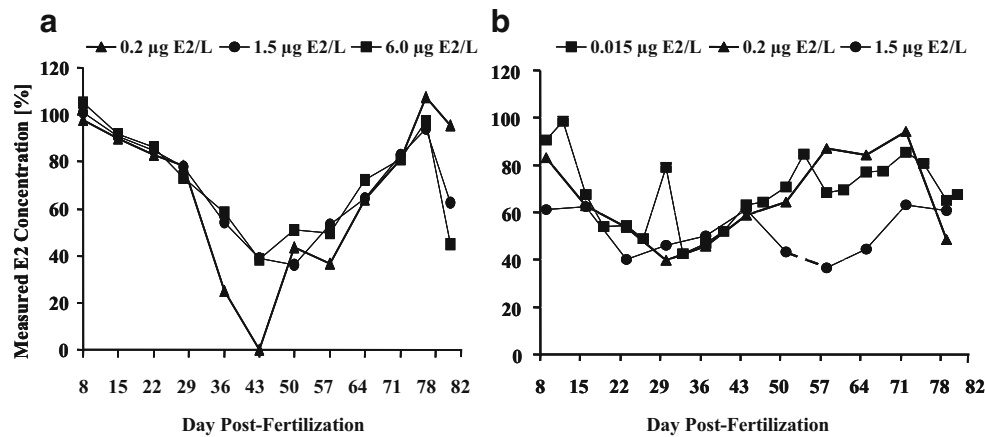


Fig. 1 Comparison of measured mean E2 concentrations in water samples of E2 treatments in the flow-through system. Mean values represent the concentrations of E2 of weekly sampling intervals from dpf 6 or 8 through the last sampling interval on termination of the experiments on dpf 82. **a**, E2 concentrations of Experiment 1A; **b**, E2

concentrations of Experiment 1B. The limit of quantification was $0.00500 \mu\text{g E2 L}^{-1}$, calculated as the product of the concentration of the lowest calibration standard ($0.100 \mu\text{g E2 L}^{-1}$) and the dilution factor of the matrix blank samples ($0.0500 \mu\text{g E2 L}^{-1}$)

rates were 100% in the negative control and 97.5, 97.5, and 99.5% in the 0.015, 0.2, and $1.5 \mu\text{g E2 L}^{-1}$ exposure groups, respectively.

Snout-to-vent length and weight

In Experiment 1A, for females the average length values were, 22.2 ± 0.1 , 22.1 ± 0.1 , and 22.8 ± 0.1 mm in the 0.2, 1.5, and $6.0 \mu\text{g E2 L}^{-1}$ treatment group compared with 22.0 ± 0.2 mm in the combined control group. Males treated with E2 had an average snout-to-vent length of 21.8 ± 0.1 , 22.1 ± 0.6 , and 23.0 ± 0.1 mm in the combined control, 1.5, and $6.0 \mu\text{g E2 L}^{-1}$ treatment group, respectively. The increase in the mean weight of males was more pronounced than the increase in the mean weight of females (Fig. 2).

In Experiment 1B no statistically significant differences between the snout-to-vent length for females and males were observed. The snout-to-vent length for females and males in

the negative control group was 17.2 ± 1.5 and 17.0 ± 1.6 mm, respectively. For females and males treated with $1.5 \mu\text{g E2 L}^{-1}$ snout-to-vent length was 17.5 ± 1.5 and 16.3 ± 1.3 mm, respectively. No statistically significant treatment-related effect was observed for body weight either in females or in males. At completion of metamorphosis, frogs from Experiment 1A weighed more than twice as much as frogs from Experiment 1B, regardless of the gender (Fig. 2).

Time course of metamorphosis

On average, the process of metamorphosis for tadpoles took longer in Experiment 1A than in Experiment 1B. In Experiment 1A, the age at which the first frog in the experiment completed metamorphosis was delayed by 6 days (dpf 44) in each E2 treatment group compared with animals in Experiment 1B (dpf 38) (Fig. 3). In Experiment 1A exposure of females to 0.2, 1.5, and $6.0 \mu\text{g E2 L}^{-1}$

Fig. 2 Body weight (means \pm SD) of **a**, females, and **b**, males, after exposure to E2 from dpf 6 or 8 in a flow-through system until completion of metamorphosis or day 82 post-fertilization. $0.015 \mu\text{g E2 L}^{-1}$ was not used in Experiment 1A and $6.0 \mu\text{g E2 L}^{-1}$ was not used in Experiment 1B. *Neg Ctrl*, negative control group

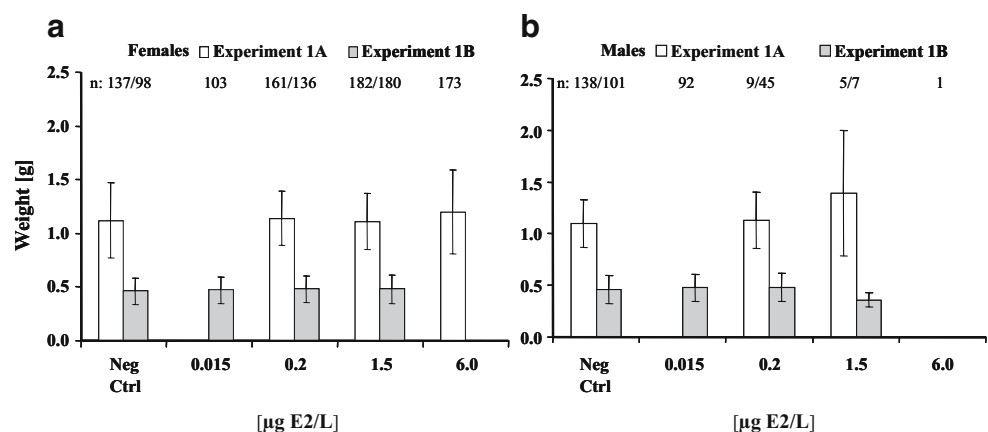
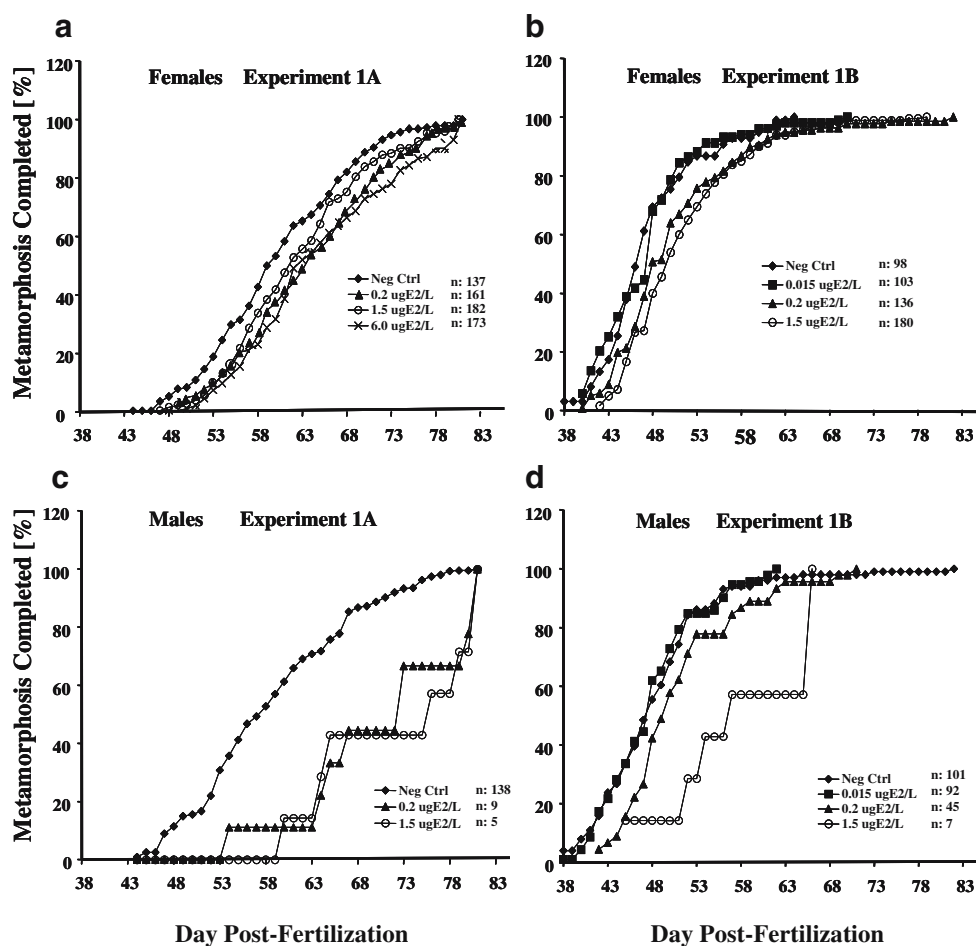


Fig. 3 Cumulative portion of *X. laevis* frogs completing metamorphosis after exposure to different concentrations of E2 from day 6 or 8 post-fertilization in a flow-through system until completion of metamorphosis or day 82 post-fertilization. The number of individuals per treatment group (*n*) is given in the legend. **a**, portion of females in Experiment 1A; **b**, portion of females in Experiment 1B; **c**, portion of males in Experiment 1A (just one male was observed in the 6.0 µg E2 L⁻¹ group); **d**, portion of males in Experiment 1B. *Neg Ctrl*, negative control group



resulted in a slight delay of metamorphosis. In Experiment 1B, the time course of metamorphosis for females in the 0.015 µg E2 L⁻¹ treatment group was similar to that for females in the negative control group but exposure to 0.2 and 1.5 µg E2 L⁻¹ caused a delay of completion of metamorphosis (Fig. 3a,b). A comparable picture for the cumulative time course of metamorphosis of males is shown in Fig. 3c,d. In Experiment 1A, males treated with 0.2 and 1.5 µg E2 L⁻¹ had a delay in completing metamorphosis of approximately ten days compared with males in the negative control group. In Experiment 1B metamorphosis of male frogs in the 0.2 µg E2 L⁻¹ treatment group was slightly postponed whereas for males treated with 1.5 µg E2 L⁻¹ completion of metamorphosis was delayed by three days. The lowest E2 concentration used obviously did not affect the mean time to complete metamorphosis.

Exposure to E2 had no effect upon the percentage of surviving tadpoles that completed metamorphosis on dpf 82 (97.4% in Experiment 1A, 99.1% in Experiment 1B). By dpf 82, 25 (out of 920) and 7 (out of 789) individuals failed to complete metamorphosis in Experiment 1A and Experiment

1B, respectively, and no treatment-related effect was observed. In Experiment 1B, 1, 2, and 4 individuals failed to complete metamorphosis by dpf 82 in the negative control, 0.2, and, 1.5 µg E2 L⁻¹ treatment groups, respectively.

Mean age at completion of metamorphosis

The differences in the time course of metamorphosis described in the previous section are reflected in the mean ages of frogs at completion of metamorphosis in each of the treatment groups. In Experiment 1A, the mean ages at completion of metamorphosis of females treated with 0.2, 1.5, and 6.0 µg E2 L⁻¹ were dpf 64.2±7.6, 62.9±7.9, and 65.1±9.2, respectively, and substantially greater than the mean age in the negative control group (dpf 61.4±8.2) (Fig. 4). In contrast, in Experiment 1B females treated with the different E2 concentrations showed a dose response with a significant increase noted in the 0.2 µg E2 L⁻¹ group (dpf 50.7±7.5) and a greater effect (dpf 51.5±6.6) in the 1.5 µg E2 L⁻¹ treatment group compared with dpf 47.8±5.6 for females in the negative control group. For males in Experiment 1A, the mean age at completion of metamor-

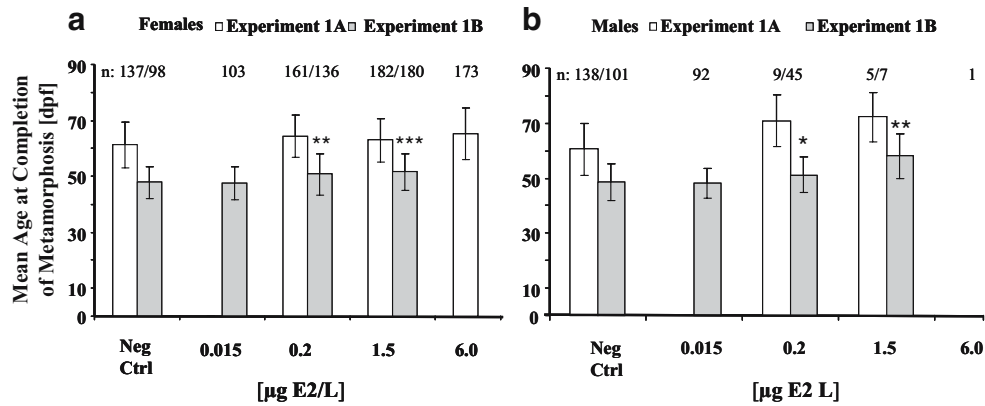


Fig. 4 Mean age at completion of metamorphosis (mean±SD) by treatment group for female, **a**, and male, **b**, frogs at completion of metamorphosis or on day 82 post-fertilization. 0.015 $\mu\text{g E2 L}^{-1}$ was not used in Experiment 1A and 6.0 $\mu\text{g E2 L}^{-1}$ was not used in Experiment 1B. Only one male frog was observed in the 6.0 $\mu\text{g E2 L}^{-1}$ group in Experiment 1A. *Neg Ctrl*, negative control group. For

Experiment 1B significant differences are marked by *asterisks* (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; two-sided Wilcoxon–Mann–Whitney test of equality with negative control group). A Kruskal–Wallis test of overall group differences was significant (females $p < 0.0001$, males $p = 0.0009$) and a Jonckheere–Terpstra test of trend was significant (females $p < 0.0001$, males $p = 0.0062$)

phosis was increased from dpf 60.6 ± 9.4 in the negative control to dpf 70.9 ± 9.3 and dpf 72.3 ± 9.0 in the 0.2 and 1.5 $\mu\text{g E2 L}^{-1}$ treatment groups, respectively. In Experiment 1B, the mean age at completion of metamorphosis of males was statistically significant greater in the 0.2 $\mu\text{g E2 L}^{-1}$ (dpf 51.2 ± 6.6) and 1.5 $\mu\text{g E2 L}^{-1}$ (dpf 58.0 ± 8.3) treatment groups than in the negative control group (dpf 48.5 ± 6.8) (Fig. 4).

Appearance and examination of non-gonadal organs

No externally visible morphologic abnormalities were observed in each treatment group in either experiment. Likewise, inspection of the liver, stomach, intestine, kidney for tumours, lesions or any other remarkable features had no effect.

Determination of sex, percent female, male, and mixed sex by gross examination

The percentages of male, female, and mixed-sexed animals were calculated based on frogs that completed metamorphosis (NF stage 66) among all animals. In Experiment 1A, the percentage of female phenotype was increased in all E2 treated groups. In the 0.2, 1.5, and 6.0 $\mu\text{g E2 L}^{-1}$ treatment group 88.4, 95.8, and 98.3% of frogs showed female phenotype. In Experiment 1B, the incidence of female phenotype in the E2 treated groups displayed a statistically significant distinct dose-responsive increase with a high percentage of females observed for the 0.2 (70.5%) and 1.5 $\mu\text{g E2 L}^{-1}$ (92.3%) treatment groups, respectively (Fig. 5). The E2 concentration that results in an increase of female phenotype to 75% is defined as the effective concentration (EC_{50}). Based on the feminization effect

observed in Experiment 1B the EC_{50} was calculated to be 0.12 $\mu\text{g E2}$.

In Experiment 1A, the percentage of mixed sex individuals were 6.9% (13/189), 2.1% (4/191) and 1.1% (2/176) in the 0.2, 1.5, and 6.0 $\mu\text{g E2 L}^{-1}$ treatment groups, respectively. In Experiment 1B, the incidences of mixed-sex animals were statistically significant with 6.2% (12/193) and 4.1% (8/195) in the 0.2 and 1.5 $\mu\text{g E2 L}^{-1}$ treatment groups, respectively (Fig. 5). In both experiments the percentage of frogs displaying mixed sex gonads was higher in E2-treated groups than in the negative control groups. When results from the two experiments are compared, the percentage of mixed sex animals was seen to be similar in equal E2 concentrations.

Segmental aplasia was detected in Experiment 1A in one of nine male individuals out of the 0.2 $\mu\text{g E2 L}^{-1}$ treatment group. In Experiment 1B, segmental aplasia was observed in four of 101 (4.0%), nine of 92 (9.8%), eighteen of 45 (40.0%) and four of seven (57.1%) male frogs in the 0.015, 0.2, and 1.5 $\mu\text{g E2 L}^{-1}$ treatment groups, respectively. Thus, the frequency of segmental aplasia in the 0.2 and 1.5 $\mu\text{g E2 L}^{-1}$ groups was significantly elevated. Pearling was observed in two of nine and one of five male frogs the 0.2, and 1.5 $\mu\text{g E2 L}^{-1}$ groups, respectively, in Experiment 1A. In Experiment 1B, the occurrence of pearling was determined in two frogs out of 45 within the 0.2 $\mu\text{g E2 L}^{-1}$ treatment group. Intersex was not found in any of the 1593 *Xenopus* evaluated for gonad morphology in both studies. Figure 6a,b demonstrate ordinary developed ovarian and testes tissue that was assumed to be “normal gonad morphology” for gross morphological evaluation. Figure 6c to e illustrate gonads showing mixed sex, pearling, and segmental aplasia and clarify the terminology which was applied in these experi-

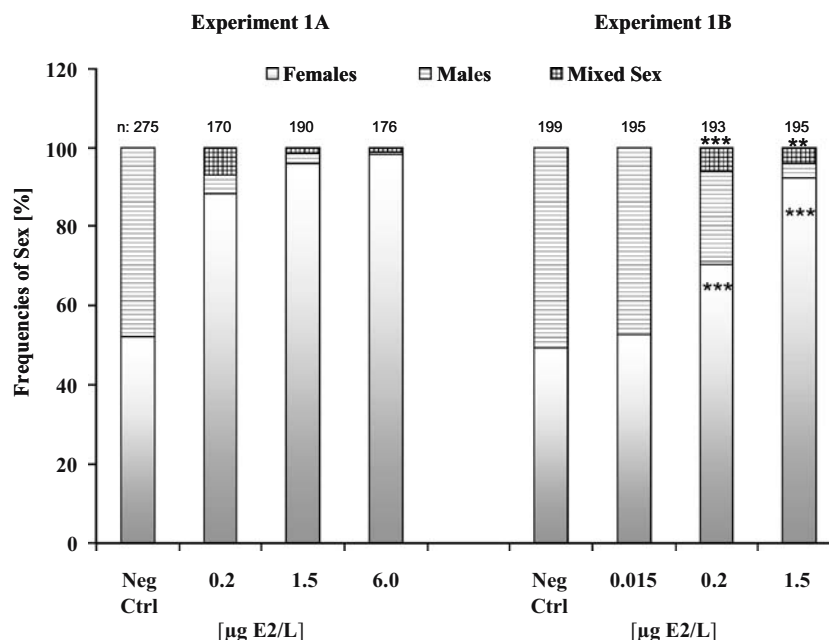


Fig. 5 Individual bars represent frequencies of sex of *X. laevis* frogs among all animals according to treatment group based on gross morphological evaluation; the total number of frogs (n) that completed metamorphosis, is given on top of the bars. Tadpoles were exposed to E2, in a flow-through system, from day 6 or 8 post-fertilization through completion of metamorphosis or day 82 post-fertilization. The EC_{50} was calculated from the results of Experiment 1B and was $0.12 \mu\text{g E2}$. The numbers of frogs in the negative control in Experiment 1A was higher for evaluation of gender because the

animals which were excluded from calculation of weight and length were included in calculation of the portion of males, females, and mixed sex. *Neg Ctrl*, negative control group. For Experiment 1B asterisks denote significant differences (** $p < 0.01$, *** $p < 0.001$; two-sided Fisher's exact test for comparison to negative control). Both an overall exact chi-square test of equality of all experimental groups and a Cochran–Armitage trend test were statistically significant ($p < 0.05$) in every case

ments to describe morphological changes in gonadal appearance of *X. laevis* due to E2 treatment. To allow the variability of the morphological appearance of gonads to be described, an extensive terminology to include the biological variability was developed and used in the two experiments (Table 1). In Experiment 1A there were apparent differences in the occurrence of ovaries that were truncated, narrow entire margin, or had enlarged masses. However, no differences in morphological characteristics were observed between E2-treated and control frogs in Experiment 1B.

Gonadal measurement

Because of the subjective nature of gross gonad morphological assessment, quantitative measurements of gonad area were performed on digital images.

In Experiment 1A, mean gonad areas of female and male frogs were approximately twice as large as in animals of the corresponding gender in Experiment 1B. The mean gonad image areas for females in Experiment 1A were $2.20 \pm 0.82 \text{ mm}^2$, $2.21 \pm 0.89 \text{ mm}^2$, $2.09 \pm 0.92 \text{ mm}^2$, and $2.23 \pm 0.92 \text{ mm}^2$ in the combined negative control, 0.2, 1.5, and $6.0 \mu\text{g E2 L}^{-1}$ treatment groups, respectively. In Experiment 1B, for females treated with 0.015, 0.2, and $1.5 \mu\text{g E2 L}^{-1}$ mean gonad image areas were $1.21 \pm 0.41 \text{ mm}^2$, $1.27 \pm$

0.44 mm^2 , and $1.32 \pm 0.41 \text{ mm}^2$, respectively, compared with $1.23 \pm 0.42 \text{ mm}^2$ in the negative control group. Mean gonad areas in females treated with $1.5 \mu\text{g E2 L}^{-1}$ were significantly increased ($p < 0.05$). In Experiment 1A, the measured mean gonad areas of males in the combined control, 0.2 and 1.5 E2 L^{-1} were $1.28 \pm 0.2 \text{ mm}^2$, $1.39 \pm 0.2 \text{ mm}^2$, and $1.70 \pm 0.3 \text{ mm}^2$, respectively. In Experiment 1B, males treated with 0.015, 0.2, and $1.5 \mu\text{g E2 L}^{-1}$ had a slight increase of mean gonad areas $0.82 \pm 0.20 \text{ mm}^2$, $0.85 \pm 0.19 \text{ mm}^2$, and $0.89 \pm 0.14 \text{ mm}^2$, respectively, compared with the negative control ($0.77 \pm 0.22 \text{ mm}^2$). A Kruskal–Wallis test of overall group differences was marginally non-significant ($p=0.079$) but a Jonckheere–Terpstra test of trend was significant ($p=0.0092$).

Discussion

Gonadal development in anuran amphibians has been proposed as a valuable model to evaluate the physiological consequences of exposure to EAC with (anti)estrogenic and/or (anti)androgenic activity [30]. It has long been known that treatment of tadpoles with gonadal steroids can disrupt normal gonadal development in several anuran species and might even lead to complete sex reversal [31,

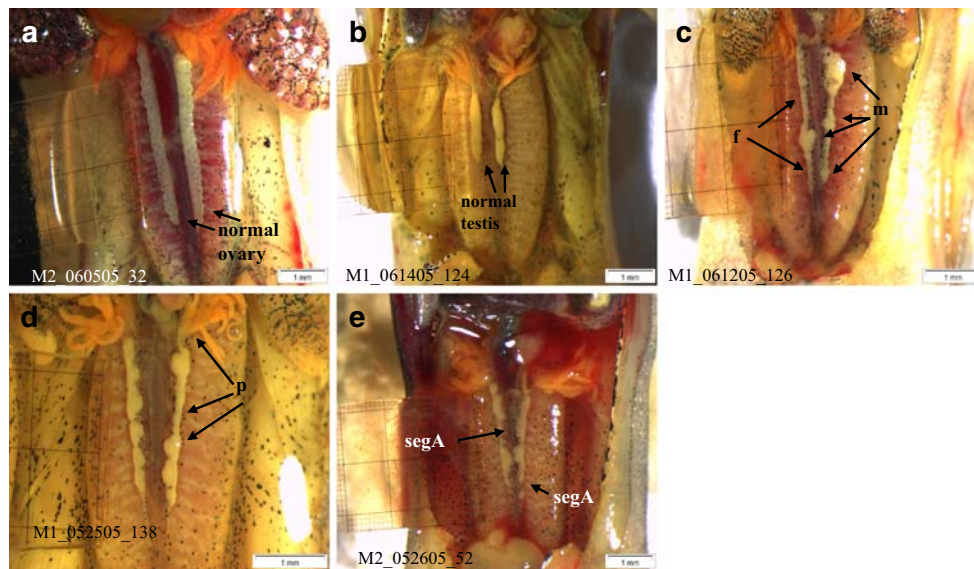


Fig. 6 **a.** Ovary of *X. laevis* revealing segmented, lobular structure and clearly visible internal melanocytes in the tissue; the length of the ovary is slightly shorter than the length of the kidney. **b.** In contrast, a testis is shorter relative to the length of the kidney (half of the kidney), straight shaped, and does not contain internal melanocytes. In general testis tissue appears much denser than ovary tissue. **c.** Mixed sex gonads were defined as the co-occurrence of both ovarian and testicular tissue in a single gonad. The ovary tissue in the presented gonad shows a reduced lobular structure, just a few melanophores,

whereas the testis tissue appears bulbous. *Arrows* indicate the different gonad structures, *f*, ovary-like tissue, and *m*, male like tissue. **d.** Pearling (*p*) was characterized by the presence of multiple, prominent segmental enlargements and/or attenuations along the length of one or both gonads. **e.** Segmental aplasia (*segA*) was classified as nodular islands of testicular or ovarian tissue, with either intervening membranous connections or no connections. The kidney and the attached gonad tissue were fixed with Bouin's solution

32]. However, studies on the potential effects of EAC on anuran gonadal development have produced variable and, in some cases, inconsistent results. From analysis of published studies it appears that some of the variability between EAC studies may be because of differences between the experimental conditions employed. One example to illustrate this situation is the outcome of three different exposure studies assessing the effects of bisphenol A (BPA) on gonadal development in *X. laevis* [15–17]. By using a semi-static exposure system, feminization of tadpoles due to BPA treatment was reported by Kloas [15] and Levy [17]. Attempts to replicate these findings in a flow-through exposure system by Pickford [16] were not successful. A more recent example includes inconsistent results observed in different studies on the putative effects of atrazine on survival, growth, metamorphosis, and sexual development of *X. laevis* [20, 21, 23]. The failure to reproduce the original findings leaves regulatory authorities with many uncertainties and in consequence, may prevent sound assessment of potential risks associated with the use of the chemicals of concern.

Contradictory results from several studies of atrazine, and the criticism associated with each of these studies, made it impossible for the US EPA to perform an unambiguous evaluation of the actual impact of atrazine on development and sexual differentiation in *X. laevis*. The

US EPA therefore requested further independent studies to be conducted using a standardized test protocol including validated positive control treatments and an optimized study design [33]. To address uncertainties resulting from inconsistencies of previously reported data, the current multi-site study was undertaken in order to develop and refine an optimized testing protocol for the use in developmental studies with *X. laevis* tadpoles. The current paper presents the results of the two independent exposure experiments that were performed by two different laboratories. A first experiment (Experiment 1A) served as a pilot study to identify appropriate testing conditions for the development of a standardized testing protocol. Accordingly, the test protocol used in the follow-up experiment (Experiment 1B) included some modifications reflecting the experience gained in Experiment 1A. In combination, the current multi-site study aimed at a characterization of the general performance of the tadpoles using the optimized testing protocol and a detailed evaluation of concentration-dependent effects of E2 on tadpole growth, larval development and sexual differentiation.

The testing protocol under investigation specified the use of a flow-through exposure system to achieve E2 exposure of tadpoles. The main advantages of flow-through systems versus static renewal systems are the maintenance of constant testing conditions (e.g., water quality) and the

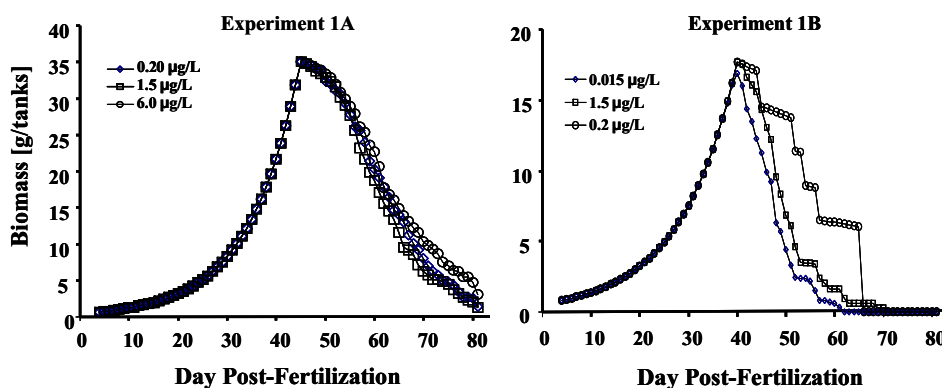
continuous delivery of the test substance to the exposure vessels to achieve stable exposure of the test species to the test substance. In the two exposure experiments with E2, however, difficulties were encountered with regard to the maintenance of the target test concentration of E2 even under flow-through conditions. Although E2 concentrations in the stock solutions were within acceptable ranges in both exposure studies, and the quantity of E2 metered into the tanks was correct, actual E2 concentrations in the exposure tanks decreased to levels well below nominal test concentrations during the entire course of the experiments (Fig. 1). These observations clearly indicate that in future studies care must be taken to maintain constant nominal concentrations of the test compound throughout exposure. It is recommended the flow rate of the stock solutions be adjusted on a daily basis according to the actual test substance concentration present in the tanks.

Comparison of the time-course of E2 concentration changes with estimates of the biomass present in E2 treatment tanks at different times during Experiments 1A and 1B suggests that E2 concentrations were inversely related to changes in total biomass. Aqueous E2 concentrations decreased concurrent with the increase in body size and body weight of tadpoles during premetamorphosis (NF stages up to 52), prometamorphosis (NF stages 53 to 57), and climax (58 to 66). Concentrations of E2 began to increase as tadpoles that completed metamorphosis were removed from the test tanks, and as remaining tadpoles began to lose weight during tail resorption. Figure 7 illustrates that E2 concentrations were inversely related to changes in total biomass of the tadpoles. Furthermore, it should be noted that the biomass was greater in Experiment 1A than in Experiment 1B and the decline in aqueous E2 concentration was more marked in Experiment 1A. These observations, combined with the relatively high octanol–water partition coefficient of E2 ($\log K_{ow}$ 3.94) suggest that additional measures (e.g., gradual increase of amounts of E2 delivered to the test tanks) are required to maintain a constant exposure to E2 in terms of actual aqueous E2 concentrations. In semi-static systems E2 disappears quick-

ly, which suggests the animal may absorb the E2. Only few studies are available that have addressed this aspect. Results from Hayes and Licht [34] suggest that exogenous E2 is readily metabolized by *X. laevis* and other anuran species, and there was no evidence of bioaccumulation of E2 in anuran tadpoles. However, metabolic enzymes and the sites of E2 metabolism in anuran tadpoles are largely unknown. Possible candidate metabolites might include various E2 conjugates, but neither the test solutions nor the test animals have been examined analytically for the presence of such metabolites. To the best of our knowledge, there are also no data available regarding possible developmental profiles of E2-metabolising capacities that might explain the transient reduction of aqueous E2 concentrations in the present studies. It is, further, not known whether any of the putative metabolites still possess estrogenic or other hormonal activity. Finally, it is possible that factors other than changes in tadpole biomass may have contributed to the observed changes of E2 concentrations. For example, a small fraction of E2 may have been absorbed either in the biofilm or directly on the surface of the glass tanks. Another possibility could be that E2 partition from water to fat in the administered Sera Micron feed (~8% fat) occurred. Notably, daily food amounts administered to each tank were adjusted to total larvae biomass. Therefore, it is an interesting possibility that at least a portion of the E2 dose may reach biological targets in the larvae as a result of oral ingestion of the feed.

An important objective underlying the development of a standardized testing protocol for exposure of *X. laevis* tadpoles was to maintain appropriate water quality throughout the exposure experiments. Thus special care was taken to standardize and refine water quality parameters to ensure optimal growth and development of *X. laevis* tadpoles. Except for a transient decline of DO concentration below desirable levels, all water-quality parameters were within acceptable limits. The decrease of DO concentrations through dpf 43 in Experiment 1A and inability to maintain the DO concentration above 60%, may have been due to insufficient cleaning. Based on this observation, the cleaning

Fig. 7 Correlation of changes in E2 concentration and the biomass of growing tadpoles during the exposure period in Experiments 1A and 1B



procedure was enhanced, and additional aeration was provided by gently bubbling air through glass pipettes in each tank.

An interesting finding in these two experiments was the effects on growth and development of *X. laevis* tadpoles that occurred due to slight differences in rearing procedures between Experiment 1A and Experiment 1B. Tadpoles in Experiment 1A required more time to complete metamorphosis and the metamorphic froglets (NF stage 66) were larger and heavier than those in Experiment 1B. Whereas the relationship between age and size is in accordance with established amphibian life-history models [25], the factors that are responsible for the inter-laboratory differences are not entirely clear in this case. In Experiment 1A, the photoperiod was 16 hours light:8 hours dark compared with 12 hours light:12 hours dark in Experiment 1B. The mean temperature was approximately 1 °C lower and more variable (± 1 °C) in Experiment 1A. The differences in photoperiod and water temperature are certainly prime candidates to explain the slightly slower development in Experiment 1A. On average, tadpoles in Experiment 1A consumed food 13 days longer than those in Experiment 1B. Subsequently, the longer food consumption of tadpoles in Experiment 1A led to greater weight compared with that of the froglets in Experiment 1B. Based on the results achieved in experiments 1A and 1B, and to consider the reported differences on growth and development of *X. laevis* in the literature, it is obviously important to standardize the test systems.

Results from Experiment 1B indicated that most of the tadpoles (92%) from the negative control group completed metamorphosis within a relatively short time (18 days), whereas metamorphosis was delayed (26 days) in a few tadpoles only (7.5%; 15/199). In contrast, time to metamorphosis in Experiment 1A, was much more variable, because 78.5% (216/275) of the tadpoles completed metamorphosis within twenty-two days and 21.5% (59/275) in thirteen days. Furthermore, tadpoles in Experiment 1A completed metamorphosis later than in Experiment 1B (Fig. 8). Thus, the use of the improved testing protocol in Experiment 1B was associated with a reduction in the time needed for completion of metamorphosis, which resulted in much more homogenous development of the test organisms. The latter finding is particularly important given the fact that still only few data are available upon which one could build a set of performance criteria to assess the validity of exposure experiments with *X. laevis* tadpoles. Homogenous growth and development in the negative control group is certainly one important aspect that deserves special attention. Generally, it can be proposed that homogenous rates of tadpole growth and development will reflect appropriate culture and testing conditions. Furthermore, any reduction in the inter-individual variability of growth and development will enable more sensitive assessment of the possible

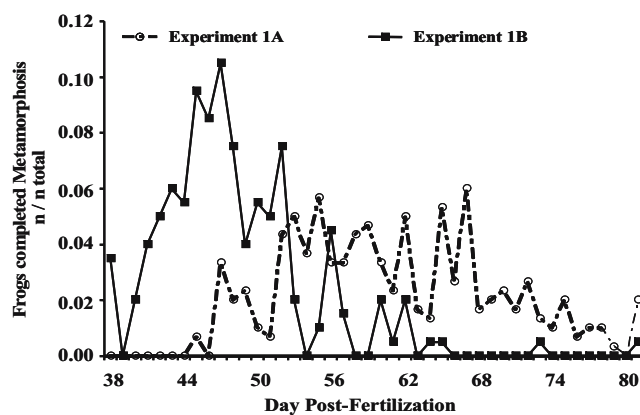


Fig. 8 Proportion of male and females frogs (n) of the negative control groups completing metamorphosis during the course of the studies, based on the total number of animals surviving in Experiments 1A and 1B

effects of a test compound on the development of tadpoles. To this end, the robust demonstration of altered rates of development due to test compound exposure will also provide additional information for interpretation of the effects of the test compound on gonadal development (e.g., identification of systemic toxicity).

In contrast, it is still difficult to assess the validity of an exposure experiment based on the time required for completion of metamorphosis. According to Nieuwkoop and Faber [25] the average time to reach stage 66 is dpf 58. In Experiment 1B, the average time to stage 66 was still shorter, being $\text{dpf } 47.8 \pm 5.6$ and 48.5 ± 6.8 for female and male frogs, respectively. This reduction of the time needed for tadpoles to complete metamorphosis is likely a result of the various efforts to optimize an array of biotic and abiotic factors known to affect tadpole growth and development.

An important element of experimental approaches used to assess the estrogenic activity of chemicals is the choice of an appropriate positive control treatment to confirm the validity of the experimental system. Therefore, a major focus of the current multi-site study was to establish a concentration–response relationship for E2 treatment on sexual differentiation of *X. laevis*. The results from Experiment 1A were not appropriate to demonstrate this relationship because all E2 concentrations (0.2 , 1.5 , and $6.0 \mu\text{g E2 L}^{-1}$) employed in Experiment 1A caused a very high percentage of phenotypic females. Based on these observations, the E2 test concentrations were modified in Experiment 1B, and the use of E2 concentrations of 0.015 , 0.2 and $1.5 \mu\text{g E2 L}^{-1}$ revealed a clear concentration-dependent increase in the number of phenotypic females. The data from Experiment 1B allowed the calculation of an EC_{50} of $0.12 \mu\text{g E2 L}^{-1}$ resulting in 75% gonadal female phenotype.

Literature information about E2 concentration-dependent effects on growth, development, and sexual differentiation in *X. laevis* is rare. In most studies single E2 concentrations

ranging from 0.3 to 100 $\mu\text{g L}^{-1}$ were used to study effects on gonadal development [15–17, 22, 31, 35, 36] and resulted in percentage of female phenotype between 60, and 100%. Hu et al. [14] applied 1, 10, and 100 $\mu\text{g E2 L}^{-1}$ and reported an increase of female phenotype to 66, 99, and 100%, respectively. Rearing *X. laevis* larvae under optimized conditions in Experiment 1B demonstrated a distinctive sensitivity for gonadal differentiation. This is demonstrated by a clear concentration-responsive effect, and a distinct effective concentration of 0.2 $\mu\text{g E2 L}^{-1}$, as shown by the frequency of female/male individuals and the occurrence of substantial mixed sex and other abnormalities. Comparing the results of Experiments 1A and 1B regarding gonadal effects of 0.2 and 1.5 $\mu\text{g E2 L}^{-1}$, a similar effect was observed, indicating high reproducibility of the testing protocol. Based on the endogenous E2 concentrations measured in adult *X. laevis* being approximately 28 ng mL^{-1} for females and 3 ng mL^{-1} for males, [37] and in froglets approximately 0.002 $\mu\text{g E2/fresh body weight}$ [38] the calculated EC_{50} appears to represent a physiological E2 concentration in *X. laevis*.

Another question of high interest is how estrogenic EAC might impact gonad gross morphology. While determination of phenotypic sex was a core endpoint in the assessment of the effects of E2 on sexual differentiation, a number of other gross abnormalities were also evaluated. To that end, a glossary of gonadal gross morphological terminology and criteria was included in the study protocol (Table 1). The importance of this glossary cannot be understated, as a substantial degree of diagnostic variability, within and among studies, can be attributed to differences in terminology [39]. For example the word intersex is characterized differently by several authors and is often used interchangeably with the term mixed sex, whereas in the glossary, these terms have different meanings. Therefore, it is difficult to describe morphological changes triggered by EAC in a way that is clear and generally accepted. Furthermore, assessments of gonadal morphology involve subjective judgement, which make studies difficult to interpret and compare. In the current multi-site study the uncertainty arising from the subjective judgement that was necessarily involved in evaluation of the gross morphological assessment each gonad was reduced by having each finding verified by a second biologist. Additionally, digital photographs of each gonad were taken to allow later verification of assessments, and to allow measurement of gonadal image area using image-processing software. Measurement of gonad image area enables objective verification of some of the subjective size-related observations made during the gross assessments. Trend analysis revealed for both sexes in Experiment 1B an increase of mean gonadal area that coincides with an increased age at completion of metamorphosis.

In both experiments, gross morphological evaluation of the gonads revealed the presence of mixed sex, segmental aplasia, and pearling at 0.2 $\mu\text{g E2 L}^{-1}$. It is noteworthy that intersex, characterized as ovarian and testicular tissue in the same individual as separate gonads (left/right), was not found in any of the 1593 frogs evaluated in these two experiments. Interestingly, in both experiments the highest incidence of these testicular abnormalities was observed in tadpoles exposed to 0.2 $\mu\text{g E2 L}^{-1}$. This was most likely due to the greater degree of complete sex reversal at the two higher E2 concentrations, as suggested by the significantly increased proportion of phenotypic females compared with phenotypic males in the 1.5 and 6.0 $\mu\text{g E2 L}^{-1}$ treatment groups. Based on gross morphological evaluation no differences for the terms of main-categories such as hyper- or hypoplasia were observed compared with individuals in the negative control in Experiment 1B.

In conclusion, this multi-site study has developed and established an optimized study design and a standardized test protocol to determine optimal growth and development of *X. laevis* larvae. High survival rates, homogenous rates of development, relatively short times to completion of metamorphosis of control animals, and concentration-dependent effects of E2 treatment on gonadal development were achieved using appropriate testing conditions. Based upon these results, it is concluded that the protocol provides a basis for evaluation on potential effects of chronic EAC exposure on gonadal development in *X. laevis* using a flow-through system. Furthermore, it is our opinion that the protocol could also be successfully applied to investigate the effects of other types of EAC, such as (anti)thyroidal chemicals, on the development of *X. laevis*.

Acknowledgement The authors gratefully acknowledge fruitful cooperation between members of Wildlife International, Ltd, IGB, EPL, Sielken and Associates, and Syngenta for providing critical and stimulating discussions. We appreciate the exceptional encouragement and assistance of all the laboratories' members involved. We also wish to thank the members of the Office of Prevention, Pesticides, and Toxic Substances from the US EPA for constructive criticism and helpful advice.

This work was funded by Syngenta Crop Protection, Inc. The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Jobling S, Nolan M, Tyler CR, Brighty G, Sumpter JP (1998) Environ Sci Technol 32:2498–2506
2. Madsen SS, Skovbolling S, Nielsen C, Korsgaard B (2004) Aquat Toxicol 68:109–120

3. Houlahan JE, Findlay CS, Schmidt BR, Meyer AH, Kuzmin SL (2000) *Nature* 404:752–755
4. Kiesecker JM, Blaustein AR, Belden LK (2001) *Nature* 410:681–684
5. Jenssen BM (2006) *Environ Health Perspect* 114(Suppl 1):76–80
6. Orlando EF, Guillette LJ (2007) *Environ Res* 104:163–173
7. Katsu Y, Lange A, Urushitani H, Ichikawa R, Paull GC, Cahill LL, Jobling S, Tyler CR, Iguchi T (2007) *Environ Sci Technol* 41:3368–3374
8. Lerner DT, Bjornsson BT, McCormick SD (2007) *Environ Toxicol Chem* 26:1433–1440
9. ASTM (2004) Standard Guide for Conducting the Frog Embryo Teratogenesis Assay-Xenopus (FETAX), E1439-98, Philadelphia, USA
10. Kikuyama S, Kawamura K, Tanaka S, Yamamoto K (1993) *Int Rev Cytol* 145:105–148
11. Shi YB (2000) *Amphibian Metamorphosis*. Wiley–Liss, New York
12. Opitz R, Hartmann S, Blank T, Braunbeck T, Lutz I, Kloas W (2006) *Toxicol Sci* 90:337–348
13. Oka T, Mitsui N, Hinago M, Miyahara M, Fujii T, Tooi O, Santo N, Urushitani H, Iguchi T, Hanaoka Y, Mikamid H (2006) *Ecotoxicol Environ Saf* 63:236–243
14. Hu F, Smith EE, Carr JA (2007) *Gen Comp. Endocrinol.*, DOI 10.1016/j.yggen.2007.04.015
15. Kloas W, Lutz I, Einspanier R (1999) *Sci Total Environ* 225:59–68
16. Pickford DB, Hetheridge MJ, Caunter JE, Hall AT, Hutchinson TH (2003) *Chemosphere* 53:223–235
17. Levy G, Lutz I, Krueger A, Kloas W (2003) *Environ Res* 94:102–111
18. Tavera-Mendoza L, Ruby S, Brousseau P, Fournier M, Cyr D, Marcogliese D (2002) *Environ Toxicol Chem* 21:527–531
19. Tavera-Mendoza L, Ruby S, Brousseau P, Fournier M, Cyr D, Marcogliese D (2002) *Environ Toxicol Chem* 21:1264–1267
20. Hayes TB, Collins A, Lee M, Mendoza M, Noriega N, Stuart AA, Vonk A (2002) *Proc Natl Acad Sci USA* 99:5476–5480
21. Hayes T, Haston K, Tsui M, Hoang A, Haeffele C, Vonk A (2003) *Environ Health Perspect* 111:568–575
22. Carr JA, Gentles A, Smith EE, Goleman WL, Urquidi LJ, Thuett K, Kendall RJ, Giesy JP, Gross TS, Solomon KR, Van der Kraak GL (2003) *Environ Toxicol Chem* 22:396–405
23. Jooste AM, Du Preez LH, Carr JA, Giesy JP, Gross TS, Kendall RJ, Smith EE, Van der Kraak GL, Solomon KR (2005) *Environ Sci Technol* 39:5255–5261
24. EPA, 2003a EPA White Paper on Potential Developmental Effects of Atrazine on Amphibians. Submitted to the FIFRA Scientific Advisory Panel, June 17, 2003 (<http://www.epa.gov/scipoly/sap/2003/june/finaljune2002telconfreport.pdf>).
25. Nieuwkoop and Faber (1994) Garland Publishing, New York, NY
26. De Boever P, Demare W, Vanderperren E, Cooreman K, Bossier P, Verstraete W (2001) *Environ Health Perspect* 109:691–697
27. Thomas KV, Hurst MR, Matthiessen P, McHugh M, Smith A, Waldo MJ (2002) *Environ Toxicol Chem* 21:1456–1461
28. ASTM, American Society for Testing Materials (ASTM) (2003) Standard guide for conducting Acute toxicity tests on test material with fishes, macroinvertebrates, and amphibians. ASTM Designation E729–96, March 2003, 22 pp
29. Greenberg A (1999) Standard Methods for the Examination of Water and Wastewater. American Water Works Association. P. 1220
30. Kloas W, Lutz I (2006) *J Chromatogr A* 1130:16–27
31. Villalpando I, Merchant-Larios H (1990) *Int J Dev Biol* 34:281–285
32. Witschi E, Allison J (1950) *Anat Rec* 108:589–590
33. EPA, FIFRA SAP. SAP Report No. 2003–01. Report:FIFRA Scientific Advisory Panel Meeting. June 17–20, 2003b, held at the Crowne Plaza Hotel, Arlington, Virginia. A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding: Potential Developmental Effects of Atrazine on Amphibians. (<http://www.epa.gov/scipoly/sap/2003/june/junemeetingreport.pdf>)
34. Hayes TB, Licht P (1993) *Gen Comp Endocrinol* 91:250–258
35. Coady KK, Murphy MB, Villeneuve DL, Hecker M, Jones PD, Carr JA, Solomon KR, Smith EE, Van der Kraak GL, Kendall RJ, Giesy JP (2005) *Ecotoxicol Environ Saf* 62:160–173
36. Cong L, Qin Z-F, Jing X-N, Yang L, Zhou J-M, Xu X-B (2006) *Aquat Toxicol* 77:250–256
37. Hecker M, Giesy JP, Jones PD, Jooste AM, Carr JA, Solomon KR, Smith EE, Van der Kraak GL, Kendall RJ, du Preez L (2004) *Environ Toxicol Chem* 1996–2007
38. Bogi C, Levy G, Lutz I, Kloas W (2002) *Comp Biochem Physiol B* 133:559–570
39. Hecker M, Murphy MB, Coady KK, Villeneuve DL, Jones PD, Carr JA, Solomon KR, Smith EE, Van der Kraak GL, Gross T, du Preez L, Kendall RJ, Giesy, JP (2006) *Rev Environ Contam Toxicol* 103–131