



EPA Public Access

Author manuscript

Environ Toxicol Chem. Author manuscript; available in PMC 2019 August 05.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Environ Toxicol Chem. 2017 December ; 36(12): 3387–3403. doi:10.1002/etc.3923.

Summary of the Development the US Environmental Protection Agency's Medaka Extended One Generation Reproduction Test (MEOGRT) Using Data from Nine Multigenerational Medaka Tests

Kevin Flynn^{*,†}, Doug Lothenbach[†], Frank Whiteman[†], Dean Hammermeister[†], Leslie W. Touart[‡], Joe Swintek[§], Norihisa Tatarazako^{||}, Yuta Onishi[#], Taisen Iguchi^{††}, Rodney Johnson[†]

[†]US Environmental Protection Agency, Mid-Continent Ecology Division, Duluth, MN, USA

[‡]Equiparent Consulting, Woodbridge, VA, USA

[§]Badger Technical Services, Duluth, MN, USA

^{||}National Institute for Environmental Studies, Tsukuba, Japan

[#]Institute of Environmental Ecology, IDEA Consultants Inc., Shizuoka, Japan

^{††}National Institute for Basic Biology, Okazaki, Japan

Abstract

In response to various legislative mandates the United States Environmental Protection Agency (USEPA) formed its Endocrine Disruptor Screening Program (EDSP), which in turn, implemented a tiered testing strategy to determine the potential of pesticides, commercial chemicals, and environmental contaminants to disrupt the endocrine system. The first tier of tests is intended to detect the potential for endocrine disruption mediated through estrogen, androgen or thyroid pathways, while the second tier is intended to further characterize the effects on these pathways and to establish a dose-response relationship for adverse effects. One of these Tier 2 tests, the Medaka Extended One Generation Reproduction Test (MEOGRT) was developed by the USEPA for the EDSP and, in collaboration with the Japanese Ministry of the Environment, for the Organisation for Economic Co-operation and Development's (OECD) Guidelines for the Testing of Chemicals. The MEOGRT protocol was iteratively modified based upon knowledge gained after successfully completing nine tests with variations in test protocols. The current manuscript describes both the final MEOGRT protocol that has been published by the USEPA and the OECD, the iterations of the protocol that provided valuable insights into nuances of the protocol, and the datasets and summaries from the various tests that informed the decision making process. The various tests include exposure to 17 β -estradiol, 4-*t*-octylphenol, *o,p'*-dichlorodiphenyltrichloroethane, 4-chloro-3-methylphenol, tamoxifen, 17 β -trenbolone, vinclozolin, and prochloraz.

* Address correspondence to Kevin Flynn, USEPA MED, 6201 Congdon Blvd, Duluth, MN 55804, (218) 529-5120, flynn.kevin@epa.gov.

DATA AVAILABILITY

Data are available at <https://catalog.data.gov/dataset/epa-sciencehub>.

Keywords

Endocrine disrupting chemical; MEOGRT; multigenerational

INTRODUCTION

The United States Environmental Protection Agency (USEPA) formed its Endocrine Disruptor Screening Program (EDSP), following the passage of several Congressional Acts in 1996, including the Food Quality Protection Act (FQPA), which amended both the Federal Food Drug and Cosmetic Act (FFDCA), and the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), and again in separate amendments to the Safe Drinking Water Act (SDWA). The EDSP has developed the requirements for the prioritization, screening and testing of pesticides, commercial chemicals, and environmental contaminants for their potential to impact the endocrine system, specifically in relation to estrogen, androgen, and thyroid hormones (<https://www.epa.gov/endocrine-disruption>; for a summary see [1]). To achieve this, the EDSP implemented a two-tiered testing strategy. Tier 1, intended to detect the potential for a chemical to have estrogen-, androgen-, or thyroid-mediated effects by various modes of action, consists of a battery of 11 validated assays. The results of the Tier 1 battery and any other scientifically relevant information are intended to be evaluated in a “weight of evidence” approach to determine, first, if a chemical indeed has the potential to interact with the estrogen, androgen, or thyroid endocrine pathways, and then next, to determine whether Tier 2 testing is needed [2]. The purpose of Tier 2 testing is to use *in vivo* testing to further characterize the estrogen, androgen, or thyroid pathway-related effects, and to establish a dose-response relationship for any potential adverse effects for risk assessment. While mechanistic endpoints are evaluated in each of the Tier 2 tests to characterize endocrine effects, it is recognized that these endpoints are not wholly independent of, and can still be influenced by, systemic toxicity. Therefore, the data from Tier 2 test are interpreted *in toto* with each endpoint contributing to the analysis.

The Tier 2 protocols cross taxa and include tests with rodents (rats), fishes (Japanese medaka, *Oryzias latipes*) and birds (Japanese quail, *Coturnix japonica*) involving exposures in more than one generation, and a long term test with amphibians (*Xenopus laevis*). For each test, a substantial validation effort, as defined by the U.S. Interagency Coordinating Committee for the Validation of Alternative Methods and the Organisation for Economic Co-operation and Development (OECD) was completed [3, 4]; the tests were presented to the FIFRA Scientific Advisory Panel of the USEPA [5] and are now accepted test guidelines within the USEPA EDSP [6] and OECD Test Guidelines Program [7].

Tier 2 tests, as the final phase of the screening and testing program, provide the most detailed information regarding potential adverse effects from the endocrine disruption activity of a tested chemical. As such, Tier 2 tests are much longer-term studies that include exposure during critical life stages and have a broad range of more tightly spaced treatments than typically used in Tier 1 screening assays. In addition, the effects associated with an endocrine disrupting chemical (EDC) may be latent and not manifested until later in life or may not be apparent until reproductive processes occur in the life history of an organism.

Therefore, the Tier 2 tests implemented in the EDSP can encompass multiple generations covering effects on fecundity and fertility, development, growth, and the transformation from juvenile life stages to sexual maturity. Successful completion of a Tier 2 EDSP test provides information to establish exposure and effects relationships, assess relevant endpoints across most life stages.

The Office of Research and Development (ORD) within the USEPA developed a test suitable for Tier 2 called the Medaka Extended One Generation Reproduction Test (MEOGRT). The Japanese medaka was chosen because of its many advantages as a test species including a relatively short generation interval (2 to 3 months), daily spawning of 25 eggs or more, relative ease of rearing including tolerance to low dissolved oxygen and a wide range of temperatures and salinity, and sexual dimorphism [8, 9] that allows comparison of gender phenotypes to the genetic sex on an individual basis. The use of medaka in laboratory testing has been detailed in several review papers [10–13], and full-life cycle tests have been developed and implemented previously [14, 15]. The MEOGRT protocol was conceived, built, and refined based upon this large body of experimental data, especially data from the medaka full-life cycle tests.

Endocrine active substances are known to disrupt endocrine regulated organizational as well as activational processes in an organism [1, 16]. Effects on organizational processes are reflected in changes to the development of the gonad or accessory tissues such that reproduction is adversely affected. The developmental changes to the gonad are often manifested as intersex gonads either with oocytes within testes [10, 17, 18] or testicular tissue within ovaries [19, 20]. However, EDCs can also cause the malformation of accessory structures essential for reproduction (ex. oviducts) that would adversely impact reproduction [21]. On the other hand, effects on activational processes result in impaired reproduction, not from structural changes to the gonad, but rather from changes in homeostatic conditions within the individual that result in changes in behavior or organ function [22, 23]. The distinction between activational effects on sexually differentiated fishes exposed to EDCs compared to the organizational effects that occurs during sexual differentiation has been generally reviewed [24]. The effects on the organization, therefore the function, of the gonad during development may not fully manifest until a later life stage. For instance, as seen in the studies presented from estrogen receptor agonists, despite rather severe intersex gonads, male medaka are still capable of successful reproduction. However, dysmorphism of the testicular efferent duct, regardless of the presence or absence of testicular-oocytes, would be expected to prohibit successful reproduction. The MEOGRT is intended to evaluate both of these types of effect. Activational effects are assessed in the test protocol with an exposure to a suspected EDC starting with naïve adults. Organizational effects are assessed in subadult and adult fish that have been continuously exposed through development and reproductive adults. In this way, both immediate and later-in-life effects are captured by maximizing the adverse impact-potential of an endocrine active substance by exposing multiple generations, and measuring pertinent endpoints at key stages within the medaka life history.

Concurrent with the efforts of ORD to develop the MEOGRT, the USEPA and the Japan Ministry of the Environment (MOE) held a series of bilateral meetings on endocrine

disruption test method development under the umbrella of the US-Japan Agreement on Cooperation in Environmental Protection, originally signed in 1975. The US-Japan discussions for developing endocrine disruptor screening and testing methodologies have been held annually since they were first initiated in 2004. These discussions formed the basis for the sharing of both data and experiences to better inform the decision making process during the development of medaka multigenerational testing protocols. In addition, these meetings were in support of an OECD project to develop a medaka multi-generation test guideline within the Endocrine Disruptors Testing and Assessment (EDTA) Advisory Group (formerly a Task Force), a part of the OECD's Test Guidelines Program. The EDTA provides advice and direction within the OECD for developing an internationally harmonized testing strategy for the screening and testing of EDCs, taking into account the consequences of such a testing strategy on the development and validation of Test Guidelines and on the existing regulatory systems for new and existing substances. Multigenerational medaka tests were conducted with specific chemicals in both countries, and the data obtained from each test was used to modify the draft medaka multigenerational test protocols until a Test Guideline was formulated that could be proposed to the OECD Validation Management Group for Ecotoxicity Testing (VMG-eco). The data and experience gained was used to iteratively develop a protocol that, in its final form, is the MEOGRT with a version published by the EPA as 890.2200 [6] and by the OECD as TG240 [7].

The starting point for the development of what would ultimately become the MEOGRT was a research protocol encompassing 2½ generations, that evaluated reproductive fitness as an integrated measure of toxicant-effects across generations, using pairs of adult medaka. The test protocol also included the measurement of a suite of endpoints for diagnostic and quantitative evaluation of EDCs or other types of reproductive toxicants. The research protocol was a relatively long-term test (29 weeks of in-life exposure) that assessed hatch, growth, survival, gonadal development, and reproduction in 2½ generations (F₀ adults, complete F₁, and complete F₂). The research protocol was an extension of existing Fish Full Life-Cycle test developed by MOE with Japanese scientists [15, 25]. After completion of several tests with variations of the research protocol, both major modifications (*e.g.* increased replication) and minor modifications (*e.g.* choice of temperature) were incorporated into a final standardized MEOGRT protocol. The current manuscript briefly presents the initial research protocol, the iterations of the protocol with various deviations, and then presents the final MEOGRT protocol. Specific changes to the protocol will be discussed with the presentation of pertinent data that supported the decision to make changes in the final protocol. In addition, the datasets from each study that informed the decision making process are included in the Supplemental Materials.

METHODS

The timelines for each of the completed tests are shown (Figure 1 Comparative Timelines). Each test represents an iterative modification of the previous tests that was intended to improve specific aspects of the protocol. To maintain coherence during the descriptions of the iterations, the initial research protocol is presented in some detail; the intermediate protocols are described simply by the modifications to the research protocol that lead to the final MEOGRT protocol.

Research protocol

The research protocol consisted of continuously exposing 2½ generations of fish to a test chemical starting with mature adults in the first generation (F₀) and exposing for five weeks to allow the test chemical and its metabolites to reach steady-state concentrations within the fish [26]. There were six treatments; five chemical treatments and clean-water control treatment. There were six replicate aquaria in each of these treatments. The measured endpoints included hatch, growth, survival, fecundity and fertility, liver vitellogenin protein or mRNA levels, the number of anal fin papillae (a gender phenotype marker), histological evaluation of the gonad phenotype, and histopathology evaluation of kidney, liver and gonad. All of these endpoints were evaluated in the context of the genetic sex of each fish. Chemical concentrations were held as constant as possible in each treatment throughout the bioassay to allow endpoint sensitivity comparisons across and within generations. The in-life exposure duration of this research protocol was 29 weeks (Figure 2 Research Protocol Timeline). Laboratory procedures involving animals were conducted in accordance with protocols reviewed and approved by one of the following: 1) the USEPA Mid-Continent Ecology Division's (MED) Animal Care and Use Committee in compliance with Animal Welfare Act Regulations and Interagency Research Animal Committee guidelines or 2) the Japan MOE in conjunction with the Institutional Animal Care and Use Committee of either the National Institute for Environmental Studies or Chemicals Evaluation and Research Institute.

F₀ adults—Young adult breeding pairs that were approximately 69 days post fertilization (dpf) (females >250 mg; males >200 mg) were moved into the exposure system and continually exposed for the next five weeks. Each replicate aquarium housed a single breeding pair for a total of 36 breeding pairs in the whole test. The fecundity of each breeding pair was assessed for the next 5 weeks by collecting all spawned eggs from each breeding pair on Tuesdays, Wednesdays, Thursdays, and Fridays. The F₀ breeding pairs were sampled at 104 dpf for various endpoints (Figure 2 Research Protocol Timeline).

F₁ and F₂: embryos to larvae—Eggs were collected from each breeding pair of the previous generation and a maximum of 20 eggs per replicate were loaded into an egg-hatching incubator. Sufficient aeration was supplied to the incubator to keep the eggs gently suspended in the water column of the incubator. This process was repeated the next day, resulting in two incubators, each containing a maximum of 20 eggs, from each breeding pair. Each day thereafter, the larval fish that hatched were collected, counted and placed in a treatment-specific aquarium. At 21 dpf, all of the surviving larval fish from each treatment were counted, pooled and redistributed to replicate aquaria with a maximum of 10 fish per aquarium and 36 total aquaria in the entire test. The survival to 21 dpf was determined at this point, and within each treatment level, extra larvae were culled. The exposure continued as the larval fish developed into subadults, defined within the protocol as about 44 dpf.

F₁ and F₂: subadult to adult—At 42 dpf, a small sample was taken from the caudal fin of each fish to determine the genotypic sex of the individual. This information was used to randomly establish six replicate XX-XY breeding pairs per treatment. All fish not selected as breeding pairs were sampled on 44 or 45 dpf for the various endpoints outlined (Figure 2

Research Protocol Timeline). From 46 dpf until the spawning assessment, the exposure continued as the subadult pairs developed into adult breeding pairs.

F₁ and F₂: adult—The fecundity of each breeding pair was assessed for 5 weeks starting at 70 dpf by collecting all spawned eggs from each breeding pair on Tuesdays, Wednesdays, Thursdays, and Fridays of each week. The fecundity assessment was completed on 101 dpf and the adults were sacrificed for adult endpoints (Figure 2 Research Protocol Timeline) at 104 dpf.

4-tert-octylphenol (4tOP) conducted by MED—MED completed one test with 4tOP using the research protocol detailed above.

Modifications to the research protocol

4-chloro-3-methylphenol (CMP), prochloraz and 17 β -trenbolone—MED completed three tests with three different EDCs, CMP, prochloraz, and 17 β -trenbolone, which followed essentially the same protocol. The most substantial modification to the research protocol in these tests was a delay in all samplings and reproductive assessments by 7 days relative to fertilization in each of the tests. In addition, while the specified temperature in the research protocol was 26°C, the water temperature during the prochloraz test was 25°C and the temperature for both the CMP and 17 β -trenbolone tests was 26°C.

Vinclozolin—MED completed a test with vinclozolin in which the exposure of the F₀ generation was started with slightly older fish than the research protocol specified. Because of this, the F₀ reproductive assessment started approximately 3 weeks later. In addition, the reproductive assessment of both the F₀ and F₁ generations ran for 21 consecutive days instead of 4 days per week for 5 weeks. Following the reproduction assessment, the adult fish were sacrificed at 114 dpf. The samples of subadult fish were taken approximately 2 weeks later than the research protocol specified on 59 and 60 dpf. The reproductive assessment in F₁ started 7 days later and again ran for 21 consecutive days. The F₂ generation was carried to the subadult stage; all fish were sampled at 59 and 60 dpf

17 β -estradiol (E2) and o,p'-dichlorodiphenyltrichloroethane (o,p'-DDT)—MOE with the Chemicals Evaluation and Research Institute (CERI) in Japan completed two tests using identical protocols, one with E2 and the other with o,p'-DDT. In all three generations (F₀, F₁, and F₂) of these tests, the reproduction assessment was started 10 days earlier than the research protocol at 60 dpf, and it was completed at 95 dpf. Immediately following the reproductive assessment of these two tests the adults were sacrificed at 96 dpf, about 1 week earlier than the research protocol.

Tamoxifen—MOE, with the National Institute for Environmental Studies (NIES) in Japan, completed a test in which there was no assessment of reproduction in F₀; only embryos were collected. In the F₁ generation, the subadult sampling was done about 10 days later than the research protocol, and while the reproductive assessment was started at the same time as specified in the research protocol, it was done for 21 consecutive days followed by a sacrifice of the adults at 93 dpf. In F₂, the test was terminated with sacrifice of fish 21 days

later than the research protocol subadult sampling, putting the age of these fish at sacrifice between the age of subadult and adult fish in the research protocol.

4tOP conducted by NIES—MOE with NIES also completed a test with 4tOP that was similar to the test done by MED using the research protocol with a few relatively minor differences. First, there was no assessment of reproduction in F₀; only embryos were collected. Secondly, the subadult sampling in both F₁ and F₂ was done about 20 days later than the research protocol at 63 and 64 dpf. Finally, the reproductive assessment for both F₁ and F₂ was done about a week later (79 dpf) and ran for 21 consecutive days instead of 4 days a week for 5 weeks as in the research protocol.

Endpoint descriptions

Regardless of the specific protocol used, the type of data collected, in general, was the same among tests. Sometimes the timing during the medaka life history was changed with the intent of providing better quality data; at other times, samples were omitted to reduce redundancy. Some of the endpoints measured are related to endocrine effects while others are traditional ecotoxicology endpoints generally considered to result from systemic toxicity, and not necessarily from endocrine effects. The primary emphasis in these tests was on endpoints that are associated with adverse effects on population-relevant parameters such as survival, gross development, growth and the reproductive measures of fecundity and fertility. To help differentiate endocrine-mediated effects from systemic toxicity, additional endocrine-mediated effects such as liver vitellogenin levels, the number of papillae on the anal fin (an external sexual phenotype marker), and the histological evaluation of gonadal sex and intersex were made. Finally, histopathology of kidney, liver, and other tissues was also assessed to better understand the relationship between endocrine-mediated responses and possible systemic toxicity. All endpoints were evaluated in the context of the genetic sex of each fish. Also, traditional ecotoxicology endpoints that may not be endocrine-mediated such as embryo hatch, growth, and survival were measured.

Dmy—Medaka have a sex-determining gene, *dmy*, that resides on the Y chromosome within a XX/XY sex determination system [27]. Therefore, individuals that have DMY are XY and under normal conditions express the male phenotype; individuals that lack DMY are XX and express the female phenotype. The genetic sex of each fish was determined by non-destructively obtaining a small tissue sample, usually from the tail fin, and using PCR methods, identifying the presence or absence of DMY [28]. All data were analyzed in the context of the individual's determined genotype not their observed phenotype. Fecundity and fertility were always measured in a XX/XY pair regardless of their external phenotype.

Growth measurements—Wet weight, measured after quickly wicking excess water from the fish, and total length, measured from the tip of the snout to the end of the tail fin, of each fish was recorded at the subadult and adult life-stages.

Liver vitellogenin (mRNA or protein)—Vitellogenin, a pre-yolk protein typically expressed in females in response to estrogen and used as a common biomarker of EDC exposure, was measured by either determining the *vitellogenin* mRNA copy number per

nanogram (ng) of total mRNA in the liver by quantitative PCR (QPCR) or by measuring the amount of vitellogenin protein per mg of liver tissue.

Vitellogenin-1 induction was determined by quantifying *vitellogenin-1* mRNA in the liver, which was dissected at necropsy. Total RNA was extracted from medaka liver stored in *RNAlater*® using RNeasy Mini Kits (Qiagen, Valencia, CA) and quantified spectrophotometrically (ND-1000, Nanodrop Technologies, Wilmington, DE). *Vitellogenin-1* mRNA was detected and quantified via reverse transcription, quantitative PCR (RT-QPCR) utilizing Taqman chemistry with primers and probes designed using Primer Express software (Applied Biosystems, Foster City, CA) and synthesized by Integrated DNA Technologies (Coralville, IA). The primer set for *vitellogenin-1*: forward (5'-AGGCAGTTTCTAAGGGCGAAC-3')/reverse (5'-TGAATGGGCATAATCTTTGTGATT-3') was designed to produce a 108 bp amplicon of the medaka *vitellogenin-1* gene (NCBI Gene ID: 100049227) with the corresponding probe (5'-FAM-TTTGGGAAATGCAAGACACCCTA-Black Hole Quencher-3') straddling an exon-exon boundary. The probe was purchased from Biosearch Technologies (Novato, CA). Characterized RNA was used to generate a standard curve to convert C_t values to copies per ng total RNA. The RNA was generated with a set of primers that produced an amplicon containing the *vitellogenin-1* amplicon that was minimally larger from total cDNA from medaka livers. This amplicon was then converted to the appropriate RNA using a T7-MEGAscript kit (Ambion, Inc, Austin, TX) and purified with a RNeasy Micro kit (Qiagen, Valencia, CA). Finally, quantity and quality of RNA product was determined on an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA).

The RT-QPCR master mix consisted of 1x Taqman buffer, manganese acetate (3 mM), dATP (300 μM), dGTP (300 μM), dCTP (300 μM), dUTP(300 μM), forward primer (500 nM), reverse primer (500 nM), probe (333 nM), rTth polymerase (0.1 U/μl), and uracil-N-glycosylase (0.01 U/μl). A thermal protocol of 1 cycle of 2 minutes at 50°C, 30 minutes at 60°C, 5 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C, 1 minute at 60°C was run with the appropriate PCR controls, standards, and samples in duplicate. The standard series covered a 6 log interval and the standard curves had a correlation of 0.98 or better. The amplification efficiency was greater than or equal to 90% of the theoretical value. Each sample was reported as the number of copies of *vitellogenin-1* mRNA per ng of total RNA.

Alternatively, the amount of vitellogenin protein per mg of liver was determined by enzyme-linked immunosorbent assay (ELISA) using commercially available kits (example, TransGenic, Kumamoto and EnBio Tec Lab, Tokyo, Japan) as previously described [29].

Secondary sex characters (SSC)—Anal fin papillae were counted with the aid of a dissecting microscope, and recorded as the SSC measurement [30].

Histopathology—Histopathology was performed on paraffin embedded fish following previously described strategies [6, 31]. Routinely, gonads, kidney, and liver were evaluated from each specimen. However, all available tissues and organs were observed and any pathologies were recorded for statistical analysis. *Chemistry*. Either previously published

methods or common laboratory methods were used to quantify the concentration of each test chemical (Table 1 Exposure Chemical Information).

Statistical analyses

All data were analyzed with procedures previously described in the final MEOGRT guidance document [6]. These procedures are briefly summarized below and in Figure 3 Flow Chart.

Histopathology data, reported as severity scores, were evaluated using a statistical procedure, the Rao-Scott Cochran-Armitage by Slices or RSCABS [32]. RSCABS uses a step-down Rao-Scott adjusted Cochran-Armitage trend test on each level of severity in a histopathology response. The Rao-Scott adjustment retains test-replication information; the by-slices procedure incorporates the biological expectation that severity of effect tends to increase with increasing doses or concentrations, while retaining the individual subject scores and revealing the severity of any effect found. The RSCABS procedure not only determines which treatments are statistically different (*i.e.*, have higher prevalence of pathology than controls), but it also determines at which severity score the difference occurs providing much needed context to the analysis.

Fecundity data were recorded daily for each replicate breeding pair. The replicate mean for each breeding pair was calculated then a square root transformation applied. A one-way ANOVA on the transformed replicate means was calculated followed by Dunnett contrasts.

All other biological data were first tested to determine whether it violates the assumptions of monotonicity by using linear and quadratic contrasts. If the quadratic contrast is significant and the linear contrast is not, the data were considered non-monotonic. If the data were monotonic, a Jonckheere-Terpstra on replicate medians trend test was done as previously recommended [33].

Transforms were not necessary to normalize the weight and length data. However, a log transformation was applied to the vitellogenin data; a square root transformation was applied to the SSC data (anal fin papillae); an arcsine-square root transformation was applied to the data on proportion hatching, percent survival, sex ratio, and percent fertile eggs.

The biological data from adult samples have one measurement per replicate (*i.e.* there are one XX fish and one XY fish per replicate aquarium). Therefore, a one-way ANOVA was done on the replicate means. If the assumptions of the ANOVA (normality and variance homogeneity, as assessed on the residuals of the ANOVA by Shapiro-Wilks test and Levene's test, respectively) were met, Dunnett contrasts were used to determine treatments that were different from the control. On the other hand, if the assumptions of the ANOVA were not met, then a Dunn's test was performed to determine which treatments were different from control. A similar procedure was conducted for data that were in the form of percentages (fertility, hatch, and survival).

The biological data from subadult samples had from 1 to 8 measurements per replicate, that is, there were variable numbers of individuals that contributed to the replicate mean for each genotypic sex since fish were randomly assigned to each replicate without knowledge of

their genotype. Therefore, a mixed effects ANOVA model was used followed by Dunnett contrasts, if the normality and variance homogeneity assumptions were met (on the residuals of the mixed effects ANOVA). If these assumptions were not met, then a Dunn's test was done to determine which treatments were different than control.

RESULTS AND DISCUSSION

A brief summary of the results from each test is provided in Tables 2 and 3, Summary of LOECs. It is not the intent of the authors to present or discuss in detail the specific effects of each chemical tested. In general, all of the tested chemicals have well-defined modes of action and the tests, regardless of protocol specifics, provided data that would be expected with each of the specific chemicals. Therefore, comprehensive data are provided in the supplemental files. The intent of the summary is to present enough information as to show that the protocols were sufficient to detect the anticipated effects of each chemical. In some cases, more detailed information is presented to support decisions to modify the research protocol into the final presented MEOGRT protocol.

Chemistry

The measured concentrations of test chemicals in all the presented studies, except the test conducted with CMP, met the performance criteria defined in the USEPA MEOGRT guideline [6], namely to, "maintain concentrations of the test chemical in solution within $\pm 20\%$ of the mean measured values over the entire test period."

Specifically, during the test with E2 conducted by CERI, the mean treatment concentrations across the entire duration of the test were (mean \pm standard deviation) 0.92 ± 0.11 , 2.8 ± 0.35 , 8.9 ± 1.0 , 27.9 ± 2.9 , and 84.3 ± 7.1 ng/L. Measurements were done every other week, and the mean concentrations were all 84% to 100% of the target concentrations.

During the 4tOP tested conducted at MED, the measured treatment concentrations were (mean \pm standard deviation) 6.2 ± 0.6 , 12.5 ± 1.5 , 25.1 ± 2.6 , 51.1 ± 5.6 , and 102.3 ± 8.5 $\mu\text{g/L}$. These means were derived from approximately 80 measurements distributed throughout the test, and the mean concentration of each treatment level was within 2% of the target concentration. The mean concentrations of 4tOP during the test conducted by NIES were recorded simply as nominal concentrations. However, given the chemical nature of 4tOP, the expertise of the laboratory conducting the test, and the similarity between the response data of the study conducted at the MED and the study conducted by NIES, the assumption has been made that the actual exposure concentrations were close to the target concentrations of 6.25, 12.5, 25, 50, and 100 $\mu\text{g/L}$ 4tOP.

During the study conducted by CERI, the mean test concentrations of *o,p'*-DDT were (mean \pm standard deviation) 0.03 ± 0.004 , 0.07 ± 0.01 , 0.22 ± 0.04 , 0.69 ± 0.17 , and 1.90 ± 0.43 $\mu\text{g/L}$. Measurements were done every other week, and the mean concentrations were 63%, 73%, 73%, 78%, 107% of the target concentration, respectively.

The measured concentrations of CMP during the test conducted by the USEPA-MED were (mean \pm standard deviation) 20.7 ± 4.6 , 43.3 ± 14.1 , 88.0 ± 27.7 , 165.1 ± 22.1 , and 344.7

± 74.4 $\mu\text{g/L}$. Excursions of greater than 20% of the mean occurred over several days in all the treatment levels, but were deemed not significant enough to compromise the outcome of the test.

The tamoxifen test conducted by NIES had mean concentrations (mean \pm standard deviation) 1.32 ± 0.19 , 2.51 ± 0.28 , 5.08 ± 0.65 , 10.17 ± 1.34 , and 20.44 ± 2.58 $\mu\text{g/L}$. Measurements were done every other week, and the mean concentrations were between 100% and 105% of the target concentration.

The measured concentrations of 17β -trenbolone during the test conducted by MED were (mean \pm standard deviation) 2.2 ± 0.5 , 5.1 ± 1.2 , 12.9 ± 2.7 , 31.7 ± 4.8 , and 84.2 ± 13.0 ng/L . Measurements were made weekly, and the mean concentrations were greater than 80% of the target concentrations.

In the MED conducted test with vinclozolin, the mean measured concentrations (mean \pm standard deviation) were 17.3 ± 1.6 , 33.2 ± 3.4 , 69.5 ± 6.2 , 136.3 ± 17.6 , and 253.3 ± 58.7 $\mu\text{g/L}$. Measurements were made weekly, and the mean concentrations were between 84% and 93% of the target concentrations.

In the MED conducted test with prochloraz, the mean measured concentrations (mean \pm standard deviation) were 5.3 ± 1.1 , 9.2 ± 1.7 , 17.5 ± 3.1 , 25.0 ± 3.8 , and 41.1 ± 3.4 $\mu\text{g/L}$. Measurements were made weekly, and the mean concentrations were between 84% and 93% of the target concentrations.

Estrogen receptor (ER) agonists

E2 (Table 2)—E2 is a well-characterized potent ER agonist. Fecundity was affected by exposure to E2, in both F₁ and F₂ generations with a lowest observable effect concentration (LOEC) of 28 ng/L . In F₁, fecundity was reduced by a greater degree with exposure to 84 ng/L E2 (no breeding pairs were available for the 84 ng/L treatment in F₂). On the other hand, reproduction was not affected by E2 exposure in F₀, indicating greater sensitivity of organizational vs activational effects on reproduction.

The non-apical endpoints can provide information about the potential adverse outcome pathway (AOP), a conceptual construct that portrays defensible linkages between a molecular initiating event (*e.g.*, an EDC binding to a receptor) through levels of biological organization eventually connecting to a population response [34]. These AOP-related endpoints responded in a manner and magnitude consistent with E2 exposures reported in the literature [35–39]. The LOEC, considering all endpoints equally, was 2.8 ng/L , based on SSC of F₁, XY subadult fish. This is the same LOEC reported from a fish full life-cycle test with medaka for F₁ [39]. In further support of the ER agonist AOP, vitellogenin production in male livers was induced by E2 exposure in the 28 ng/L and higher treatments. Testicular oocytes started to appear at 28 ng/L as well. Growth was generally not altered by E2 suggesting no systemic toxicity.

4tOP by MED or NIES (Table 2)—4tOP is a relatively strong estrogen that induces vitellogenin production and abnormal sex differentiation in medaka after exposing for 60

days (starting with embryos) at 11.4 µg/L [40]. The literature clearly indicates that 4tOP is an ER agonist [41–44]. Therefore, it is not surprising to have observed decreased reproduction in F₁ at 102 µg/L and in F₂ at 51 µg/L. At 102 µg/L, reproduction essentially ceased, while at 51 µg/L, 4tOP reduced fecundity by approximately 50%. On the other hand, reproduction was not affected by 4tOP exposure in F₀, again indicating greater sensitivity to organizational vs activational effects on reproduction.

Endpoints expected to be affected by an estrogen responded to 4tOP exposure. The overall LOEC for the test was 13 µg/L, based on SSC in F₂ XY fish, similar to published LOECs from other studies [40, 45, 46]. In support of the ER agonist AOP, the number of anal fin papillae was lowered in XY fish in both F₁ and F₂ at both subadult and adult samplings. Further, *vitellogenin* gene expression was induced in XY fish at subadult and adult lifestages in all generations with LOECs of either 25 µg/L or 51 µg/L, and the prevalence of testicular oocytes in XY fish increased at all samplings except F₀ with LOECs ranging from 13 µg/L to 51 µg/L.

Histological observations were also made to provide further information about both potential endocrine and non-endocrine effects on multiple organs including gonad, liver, and kidney. Of the pathologies reported, the majority were observed in XY individuals and could be linked to an ER agonist AOP: increased number and severity of testicular oocytes (LOEC = 13 µg/L), developmental perturbation within the gonad related to both cell types (immature spermatocysts; LOEC = 13 µg/L) and structural malformations (efferent duct; LOEC = 13 µg/L), testicular hypoplasia (*i.e.*, smaller than normal testes), and increased basophilia in the liver, presumably related to the production of vitellogenin (LOEC = 51 µg/L). There were also pathologies observed in the kidneys of XY fish that are most likely a secondary pathology induced by high concentrations of vitellogenin in the bloodstream of XY fish that normally lack a clearance mechanism for vitellogenin. There were very few, if any, pathologies that would be linked to non-EDC toxicity.

In addition to the 4tOP test completed by MED using the research protocol, NIES conducted a similar test, again with 4tOP as the test agent. Because some of the data provided by NIES were summary information with supporting statistical analyses, and not primary data, the data was not analyzed using the statistical procedures specified in the guideline.

Fecundity was reduced by about 80% in F₁ after exposure to the highest concentration tested, 100 µg/L, which resulted in no fish in that treatment in F₂. No changes in fecundity were observed in F₂ at the highest concentration, 50 µg/L. These results are very similar to the test conducted by MED where the LOEC for fecundity was 102 µg/L in F₁ and 51 µg/L in F₂.

The secondary endpoints that provide AOP information responded as would be anticipated as well. Anal fin papillae in XY fish were reduced at a LOEC of 6.25 µg/L in F₁ subadults and 25 µg/L in F₂ subadults, while vitellogenin production in XY fish was induced at a LOEC of 6.25 µg/L in F₂ and only 25 µg/L in F₁. The mean hepatic vitellogenin concentration in the same treatment across F₁ and F₂ generations was not very great, 4.8 ng/mg liver in F₁ compared to 8 ng/mg liver in F₂, but in F₂, it was high enough to be

statistically different from controls. The effect on the gonad phenotype was different between the MED and the NIES tests; many XY fish with testicular oocytes were observed in the MED test and very few were reported in the NIES study.

***o,p'*-DDT (Table 2)**—*o,p'*-DDT was chosen because it is a known bioaccumulative ER agonist [47, 48]. Reproduction was not effected by exposure in F₀, even at the highest treatment 1.9 µg/L. However, in F₁, 1.9 µg/L *o,p'*-DDT reduced fertility but not fecundity. Two of the exposed pairs produced large numbers of non-viable eggs. The LOEC for fecundity and fertility in F₂ was also 1.9 µg/L. In this treatment, fecundity was 14 eggs per pair each day compared to 20 eggs per pair each day in controls, and the mean number of fertile eggs was only 7 per pair each day compared to 18 fertile eggs per pair each day in controls. Interestingly, these means for exposed breeding pairs were similar to the reproductive output in the same 1.9 µg/L treatment of F₁; however, the control mean in F₂ was higher than in F₁, which provided greater statistical power in the data analysis.

The overall LOEC for the test was 0.07 µg/L, based on an increase in the hepatic vitellogenin concentration in adult XX fish. The magnitude of this increase was likely not biologically significant, and did not increase with higher concentrations of the test agent. The hepatic vitellogenin concentration increased in adult XY fish in both F₁ and F₂, but only in the highest treatment (*i.e.*, 1.9 µg/L).

The secondary endpoints again corroborate the expectation of *o,p'*-DDT acting as an ER agonist. The numbers of anal fin papillae were lowered by more than 50% in adult fish, and hepatic vitellogenin was increased in XY subadult fish in the 1.9 µg/L treatment. Papillae were not present in subadult XY fish but were present in the adult XY fish of the same treatment, suggesting papillae development was delayed. In addition, most of the XY fish exposed to 1.9 µg/L *o,p'*-DDT in both F₁ and F₂ generations and at both developmental stages had testicular oocytes.

CMP (Table 2)—CMP is a weak estrogen agonist as it binds to both recombinant human and rainbow trout (*Oncorhynchus mykiss*) ERs, and in rainbow trout liver slices and dissociated hepatocytes, induces vitellogenin mRNA at about the same concentrations as overt toxicity [49]. CMP has also been shown to bind to the ER in a yeast two-hybrid assay [50].

No endocrine effects were observed in our test except for possible induction of *vitellogenin* gene expression in XX adult fish in F₀, in which only a small increase (~2X of controls) was observed at 88, 165, and 345 µg/L of CMP. However, this conflicts with the observation that vitellogenin expression was decreased in subadult XX fish in the F₂ generation after exposure to 345 µg/L CMP. Also, XY subadult fish in F₁ but not F₂ were smaller (*i.e.* decreased weight and length) than control fish upon exposure to 165 and 345 µg/L CMP. Regardless, reproduction was not decreased in any of the generations. In addition, no indication of toxicity was observed via histopathology. Therefore, the medaka multigeneration study seemed to indicate that the LOEC for CMP was greater than 345 µg/L.

To confirm this, after the test, MED performed a short-term test to establish the medaka LC₅₀ to CMP and to determine whether concentrations higher than 345 µg/L cause EDC-related impacts. Newly hatched fish were exposed for 35 days to CMP at concentrations of 300, 600, 1200, 2400, and 4800 µg/L. Survival was monitored throughout the exposure and *vitellogenin* gene expression was quantified at the end of the exposure. The 7-day and 14-day LC₅₀ values were 2840 µg/L and 1090 µg/L CMP, respectively. After the 35-day exposure, only fish from control, 300 µg/L, 600 µg/L, and 1200 µg/L treatments survived and liver *vitellogenin* mRNA levels did not change in either XX or XY fish which indicated that CMP is overtly toxic at a lower concentration than required to induce EDC-related effects in medaka.

ER antagonist (Table 3)—Tamoxifen is classified as a selective ER modulator as it has multiple actions both as an ER antagonist and in specific tissues/species as an ER agonist [51]. After binding to the ER, tamoxifen has been shown to act as both agonist and antagonist in various species including the medaka, in which previous tests have shown increased plasma vitellogenin in males but decreased plasma vitellogenin in females [52, 53]. Furthermore, a relatively high dose of tamoxifen (200 µg/L) skewed the sex ratio of tilapia (*Oreochromis niloticus*) to 90% males after a 60-day exposure starting with fry [54].

In the tamoxifen test, fecundity was reduced by over 50% upon exposure to 20 µg/L, the highest concentration tested. The AOP-related endpoints were typically affected at a LOEC of 1.3 µg/L and subadults were more sensitive than adults. The affected LOEC endpoints were increased vitellogenin and decreased SSC in XY fish. Similar to other studies [52, 53], tamoxifen actually reduced vitellogenin in female fish, starting at a concentration of 1.3 µg/L. The vitellogenin induction in males was fairly modest, only a few times higher than controls, and not approaching the same vitellogenin concentration as control female fish. The maximum change in SSC was seen in F₂ where the number of anal fin papillae was reduced from about 80 per control XY fish to approximately 25 per XY fish in the 1.3 µg/L treatment. These same endpoints had slightly higher LOECs (from 1 to 2 treatment levels) in the adult life stage, due either to true differences in the sensitivity of the life-stage, or to reduced statistical power from smaller sample sizes in the adult samplings. Also note, the subadult samplings were not conducted at the same developmental stage across generations; F₁ was sampled at 54 dpf and F₂ was sampled at 65 dpf, thus, the control values for these parameters were quite different from each other.

In addition to the endocrine AOP endpoints, tamoxifen reduced growth by about 30% in both XX and XY fish at 54 dpf (F₁), but by less than 20% in only the XX fish at 65 dpf (F₂). It is unclear whether this difference is due to delayed maturation, or irreversible growth effects. Unfortunately, growth data, which would have provided insight into this question were not recorded for adults. Histopathology, which would have provided information both about potential EDC impacts and non-EDC toxicity at the organ/tissue level, was also not assessed.

Androgen receptor (AR) agonist (Table 3)—17β-Trenbolone acts as a potent agonist of AR in a variety of systems including fish, where it has been shown to reduce fecundity, vitellogenin, and generally induce masculinization of both external SSC and gonad

phenotypes [18, 19, 55–57]. It is androgen-dependent expression of key genes (*Bmp7* and *Lef1*) that is required for bone nodule outgrowth leading to the formation of papillary processes, SSC, in medaka [58]. In a previous 21-day exposure of medaka, the LOEC for changes in SSC and decreased vitellogenin was 365 ng/L and 39.7 ng/L, respectively [38].

In the MED test the LOEC for reproduction, which occurred in F_1 , was 32 ng/L, where fecundity was reduced to zero. In addition, 13 ng/L 17β -trenbolone seemed to reduce fecundity (control fecundity equaled 25 eggs per pair-day versus 11 eggs per pair-day in the treatment); however, it was not statistically significant. No reduction in reproduction was observed in F_2 where the highest concentration was 13 ng. The mean fecundity in this treatment was equal to controls (24 eggs per pair-day) and much higher than in the same treatment in F_1 (11 eggs per pair-day).

The LOECs across the entire 17β -trenbolone test in the AOP-related endpoints ranged from 32 to 84 ng/L, predominately affecting XX fish. All the findings supported the expected response to an AR agonist (*i.e.*, the presence of anal fin papillae (the measure of SSC) in XX fish, decreased *vitellogenin* gene expression in XX fish, and most dramatically, XX fish presenting complete male phenotypes including functional testes). In the 84 ng/L treatment, essentially all XX fish presented the male phenotype with respect to gonad, *vitellogenin* gene copy number, and SSC. In addition, in the 32 ng/L treatment the same effects were observed at intermediate magnitudes except for the gonad phenotype in which, interestingly, intersex gonads were never observed, only complete sex reversal. However, intermediate effects in gonad phenotype were observed in a slightly different manner. For instance, in F_1 subadults, 16 of 16 XX fish exposed to 84 ng/L 17β -trenbolone had testes, while in the 32 ng/L treatment, only 7 of 17 XX fish had testes.

The observed pathologies further supported the androgen receptor AOP. First, there were very few pathologies found in XY fish. Second, many of the pathologies observed in XX fish could be ascribed to an AR agonist exposure. For instance, oocyte atresia and decreased yolk formation were found in XX subadults and adults at mild to moderate severities. There were also pathologies observed in the kidneys of XX fish at high prevalence and severity including tubule epithelial hypertrophy and eosinophilia. Ironically, in the 32 ng/L treatment and higher, 17β -trenbolone, completely caused sex reversal in XX individuals. In spite of the fact that they were XX fish, these fish had histologically normal testes.

In general, the F_0 generation was less sensitive than the F_1 . Because there were no viable fish available in the 32 and 84 ng/L treatments of F_2 , no effects were observed in any of the measurements of this generation. In addition, the LOECs obtained at the F_1 subadult sampling were lower than those of the F_1 adult sampling. However, it cannot be determined whether this was a result of a greater sensitivity at this life stage or due to the higher statistical power from larger sample sizes per replicate at the subadult sampling.

The suite of biological impacts observed across levels of biological organization were consistent with the well documented AOP of 17β -trenbolone as an AR agonist. At the population/organismal level, fecundity was reduced starting at 32 ng/L, reflecting the 17β -trenbolone effect on the phenotype of individual XX medaka. Starting at the 32 ng/L level,

XX fish developed into phenotypic males across multiple levels of organization including molecular (vitellogenin), tissue (external SSCs and pathology), and organ (gonad phenotype), all tethered back to the molecular initiating event (*i.e.*, 17β -trenbolone binding and activating the AR). *AR antagonist* (Table 3). Vinclozolin and its degradants are anti-androgens that persist in the environment for over 180 days [59, 60]. The presumptive AOP initiated by vinclozolin is binding with the AR without initiating the transcription of androgen-dependent genes [61, 62].

In F_0 , fecundity was decreased in the 253 $\mu\text{g/L}$ treatment starting in the second week of the reproduction assessment, the third week after the exposure started. The mean for the entire three-week reproduction assessment was just 7 eggs spawned per day in the 253 $\mu\text{g/L}$ treatment, compared to 24 spawned eggs per day in controls. Significant, but not large reductions in fecundity were also seen in all the other treatment groups (17, 33, 70, and 136 $\mu\text{g/L}$) in the last week of the assessment; control fecundity was 24 eggs per day and all the treatments spawned approximately 15 eggs per day. For comparison, in a 21-day test measuring reproduction in fathead minnows (*Pimephales promelas*), vinclozolin reduced cumulative egg production by about a third at 60 $\mu\text{g/L}$ [61], while in medaka, vinclozolin at 640 $\mu\text{g/L}$ did not reduce the number of eggs produced in a 21 day assessment following OECD TG 229 [63].

In F_1 of the current assay, no vinclozolin effects were observed at the concentrations tested. The mean fecundities were 25, 26, 20, 29, and 26 eggs per pair per day in the control, 17, 33, 70, and 136 $\mu\text{g/L}$ treatments respectively. There was almost no reproduction in the 253 $\mu\text{g/L}$ treatment; however because there was only a single breeding pair in this treatment the fecundity was not statistically different from controls.

There was little evidence to link vinclozolin to an anti-androgen AOP. *Vitellogenin* gene expression was increased in XY subadults in only the highest treatment in F_2 , 136 $\mu\text{g/L}$. In this group, very high measurements were observed in a few individuals which accounted for the statistical significance. Contrary to these observations, in a short-term exposure with medaka, hepatic vitellogenin levels in adult females were reduced at 640 $\mu\text{g/L}$, but not 200 $\mu\text{g/L}$ vinclozolin [63]. Possibly linked to an anti-androgen AOP, anal fin papillae were consistently decreased in XY subadults with a LOEC of 17 $\mu\text{g/L}$ in F_1 and 33 $\mu\text{g/L}$ in F_2 . In these treatments, the anal fin papillae were reduced by 25% and 40%, respectively, which was similar to the effects of vinclozolin had on juvenile medaka exposed for only 28 days [63]. In that study, the effects on anal fin papillae were seen with exposure of 640 $\mu\text{g/L}$ which caused a reduction of slightly over 50%. In our study, vinclozolin exposure caused only a delay in SSC development, as the number of papillae counted in exposed adults in both F_1 and F_2 were similar to control adult values.

While growth in subadults was not affected by vinclozolin, it did increase the weight and length of adults in some treatments. In F_0 , XY fish were heavier, 275 mg in controls versus 350 mg in the 33 $\mu\text{g/L}$ treatment; however, the weight did not increase with higher concentrations of vinclozolin, and XY adult fish were not affected in F_1 . In addition, XX adult fish were heavier in both F_0 and F_1 .

There were few, by-and-large low severity pathologies observed, and even fewer that were consistently observed across generations. The most prominent pathology was eosinophilia of the tubular epithelium in the kidneys of F₁ subadult XY fish. The LOEC of this pathology at a moderate severity (3) was 17 µg/L. In F₂, the LOEC for the same pathology at the same severity was 70 µg/L. There was no evidence of vinclozolin-induced change to the gonad gender phenotype.

Vinclozolin rapidly hydrolyzes in water to more potent anti-androgens than the parent chemical [64, 65]. Because of this, the characteristics of the exposure system delivering toxicant to the aquaria (*i.e.*, flow rates, residence time of toxicant, *etc.*) can quantitatively change the relative concentrations of the exposure mixture, comprised of the parent chemical and the daughter hydrolytes. The complexity of the exposure system makes it difficult to compare the results from different vinclozolin aquatic bioassays, especially if different exposure infrastructures are used. Whether this explains why results differ across the limited number of vinclozolin studies with medaka remains to be seen.

There were two treatments with XY fish that expressed normal female phenotypes, (*i.e.*, normal ovaries, no anal fin papillae, and normal vitellogenin levels); three XY fish in 33 µg/L in F₁, and four in 136 µg/L in F₂. These outcomes were not seen in any other treatments including controls. Two explanations for these observations include, either DNA contamination of XX samples from XY fish during the process of determining genetic sex, or the data represents a true response to vinclozolin exposure. It seems unlikely that this response was caused by vinclozolin because the complete dataset does not correspond with results caused by an EDC with a strong gonadal effect. For instance, there were no intersex gonads observed in XY fish from either higher or lower treatments. Additionally, no suite of pathologies was observed in the gonad or liver relatable to an EDC exposure.

Steroidogenesis inhibitor (Table 3)—Prochloraz is a fungicide that inhibits sterol biosynthesis in fungi. Because of the conserved nature of the pathways involved, prochloraz also inhibits CYP19 aromatase in fish, which catalyzes the conversion of testosterone to E2, both in the ovary and the brain [66–68]. There is also evidence that prochloraz acts as an AR antagonist. In a 21-day fathead minnow reproductive assay, prochloraz produced a suite of biological effects characteristic of an aromatase inhibitor and an AR antagonist including alterations in various steroid levels, reduced vitellogenin in females, and lower fecundity at 0.1 mg/L prochloraz [69]. In addition, fish embryo toxicity tests using medaka and zebrafish showed that prochloraz effected gene transcription as predicted for an anti-androgen [70]

In the current assay, the LOEC for fecundity was 41 µg/L in F₀ (the highest concentration tested) and 25 µg/L in F₂. No LOEC was determined up to the highest test concentration in F₁, (25 µg/L). The mean fecundities were similar in the 25 µg/L treatments of F₁ and F₂: 4- and 6 eggs per pair-day, respectively. However, due to lower than normal fecundity in the F₁ control breeding pairs (three of the six breeding pairs in F₁ did not spawn), only the 25 µg/L treatment in F₂ was significantly different from control. The increased number of replicates specified in the final MEOGRT guidance will mitigate the reduction in statistical power caused by random non-spawning breeding pairs.

The LOECs of the other endpoints ranged from 5 to 41 µg/L, affecting both XY and XX fish. Prochloraz affected growth in relatively low concentrations but the decreases observed in the subadult samples were not very large and were not observed in adults. In fact, exposed adults in F₀ were actually larger than controls. In addition, there were decreased copies of the *vitellogenin* gene in XX fish. This can be plausibly linked to aromatase inhibition, causing circulating E2 to decrease, which in turn would lower the expression of hepatic vitellogenin. Lastly, there was a modest reduction of anal fin papillae in F₁ subadult XY fish, but this was not observed in F₂. Interestingly, this was not observed in adults in any of the three generations.

Many of the pathologies observed in the prochloraz test were not present consistently through the three generations or across different life stages. While this confounds interpretation of the histopathology data, some generalities can be made. Prochloraz did not affect the tissues and organs of XY fish, but it did cause EDC-related pathologies in the ovaries of XX fish (e.g., decreased yolk and follicular hyperplasia/hypertrophy). As more information is gathered about steriodogenesis inhibitor induced pathologies, it may be possible to interpret histopathology datasets like this, more clearly.

Rationale for changes to the research protocol

The outcomes of the various tests presented with several different chemicals acting through differing AOPs have provided enough information to develop guidance for performing an extended one-generation medaka reproduction test for use as a fish test in Tier 2 of the EDSP. Two major changes were made to the original research protocol: 1) the number of replicates per treatment was increased to strengthen the statistical evaluation of effects on reproduction, and 2) nearly all of the F₂ generation was removed; the test was terminated after the embryos hatched in F₂. Other more minor changes included evaluating pathology in only the F₁ adults sampled after the assessment of reproduction; sampling subadult fish when slightly older, eliminating samples that were redundant, and adjusting the temperature to optimize reproduction while minimizing the occurrence of anomalous XX males. The MEOGRT Exposure and Endpoint Timeline, Figure 4, presents the timeline for exposure and collection of endpoint data for the adopted MEOGRT guideline [6, 7].

The reproductive endpoints, fecundity and fertility, are apical ecotoxicology endpoints due to their direct relationship to population-level effects. To assess treatment effects on reproduction, the research protocol employed six replicate aquaria for each test treatment, including the controls. The fecundity data from some of the earlier the tests had relatively high variance. With only six breeding replicates per treatment, the statistical power for detecting significant changes in fecundity was quite low [71]. Minor modifications of the protocol were implemented to increase the statistical power of the protocol by increasing baseline fecundity and reducing the variance in egg production. While only marginally successful, the trends observed in the fecundity dose-responses suggested that it was being affected at treatment levels lower than the statistical LOEC. In some cases, the fecundity in chemical treatments had to be reduced to nearly zero to be significantly different from the controls. To increase the statistical power of the test to detect treatment effects on reproduction, the number of replicate breeding pairs was increased.

Replication—To determine the number of replicates preferred, a power analysis was performed using the reproductive performance of control fish from several of the tests [71]. The fecundity data had several characteristics that complicated the simple straightforward calculations of statistical power. For example, the data could not be transformed into normal distributions, and, days without reproduction (*i.e.*, when the number of spawned eggs by a breeding pair is zero) were randomly distributed throughout the data set. Because of these characteristics, a simulation was used to calculate the effect of the number of replicate pairs per treatment on the power to detect a reduction in fecundity. Initially balanced designs were considered where the number of replicates were equal in all treatments including controls. Then, based upon the recommendations from the OECD [33], we investigated designs where the number of control replicates are double the number of replicates in each treatment (2:1 design). The power properties of both the balanced design with six replicates in each treatment and the 2:1 design with 12 replicates in controls and six replicates in each treatment were considered inadequate. For example, the power to detect a 40% reduction in fecundity was 0.66 in the balanced and 0.76, in the 2:1 design. However, the power properties of a 2:1 design with 24 replicates in controls and 12 replicates in each treatment are quite good (>0.80 power) to detect even a 30% reduction (Table 4). Besides statistical power, consideration was given for the possibility of mortality, especially in the control replicates. During reproduction, a small percentage of the adults die, irrespective of treatment. Additionally, even with skilled technicians, there is a possibility of handling-induced mortalities. To be conservative, 12 breeding pairs (replicates) in treatments and 24 control breeding pairs (replicates) was chosen to mitigate the consequences of these mortalities on statistical power.

Elimination of F₂ from protocol—Considerations regarding the inclusion of F₂ in the MEOGRT were based primarily on whether the additional information gained was worth the additional costs of time, resources and risk of test failure. To be included, the F₂ generation should add significant value to understanding treatment effects on the test population. Summarizing data from all the tests that included fecundity assessments in both F₁ and F₂, indicated little evidence for consistent differences between the generations. With E2 the LOEC for fecundity was lower by one treatment level in F₁ compared to F₂. In one of the two tests with 4tOP, the LOEC in F₂ was one treatment level lower and in the other test, the LOECs were the same in both generations. In the *o,p'*-DDT test, the LOEC was not determinable in F₁ because there was no effect in the highest treatment, 1.9 µg/L; in F₂ the LOEC was 1.9 µg/L. Lastly, the LOECs for fecundity in F₁ and F₂ for the 17β-trenbolone test were equal. For these five tests using three different chemicals, the LOEC was lower in F₁ in one test; lower in F₂ in two tests, and the same in two tests. In all cases, when the LOECs were different across generations, it was by one treatment level. Additionally, the mean fecundity values in the control treatments and the LOEC treatments were quite similar across generations. Finally, the mean fecundities for any particular treatment were not statistically different across generations. This provides the clearest evidence that fecundity in F₂ was no more or less affected by EDC exposure than in the F₁ generation.

Occasionally, there is a LOEC for a non-fecundity endpoint that is lower in the F₂ generation as compared to F₁. In all cases, this LOEC is different by only one treatment level.

When considering the use of multiple generations in a test protocol, it is appropriate to consider whether the information provided by adding another generation is worth the risk of not completing the test protocol successfully. Longer tests face greater the risk of incidents such as malfunctions of exposure and/or dilution pumps, temperature control systems, disease, and, perhaps the most common problem -- maintaining consistent treatment concentrations throughout the test. Heterotrophic microbes, which use the test chemical as a nutrient source, tend to thrive and proliferate in the exposure system during long tests with subsequent reductions in the treatment concentrations. Periodic sterilization and adjustment of the exposure system are necessary to mitigate this effect.

For the above reasons the decision was made to end the MEOGRT protocol after the hatch of the F₂ embryos. Under the scenarios tested thus far, continuing the F₂ generation did not add significant understanding of treatment effects. However, it is possible that future assessments could indicate the need to continue the exposure into the F₂ generation. For instance, chemicals with high bioconcentration potentials or with indications of transgenerational effects may warrant extending the test further into the F₂ generation.

Histopathology—Increasing the number of replicates to twelve in the MEOGRT compared to six in the research protocol doubles the number of adult fish for the pathology analysis. This difference is substantial. It provides pathology results for up to 12 fish per gender per treatment. Also, because the pathology is done on adult fish sampled after the reproduction evaluation, the pathology information can be correlated with the reproductive performance of each specific breeding pair. This is in contrast with subadult pathology analysis which does not allow these direct comparisons to be made. Evaluation of treatment-related effects on gonadal development such as sex-reversal, and the prevalence testicular oocytes, is possible in either adult or subadult specimens, however, only the adult samples provide data that can be used to correlate pathology with reproduction. The increased sample sizes for adult pathology in the MEOGRT increases both the statistical power and the relevance of the pathology endpoint compared with the research protocol.

Sampling time of F₁ subadult fish—In most of the tests, the subadult sampling was done, depending on the specific tests, between 44 and 59 dpf. Within this time window of the MEOGRT, medaka are growing and developing rapidly enough that shifting the subadult sampling time a week earlier or later can have dramatic impacts on the data collected. For instance, typical medaka growth curves with parameters similar to those specified in the MEOGRT protocol are steep at this time (Figure 5 Growth Curve). So while the qualitative impacts on growth by an EDC would not be changed by the sampling time, there would most likely be a quantitative impact by shifting the sampling time on data such as the mean weight or length of controls. Therefore, the MEOGRT guidelines recommend sampling subadults later in development (9 wpf) than in the research protocol (6 wpf). Interestingly, as can be seen in the growth curves (*i.e.*, length), the rate of growth of control fish is consistent across the two laboratories (MED in the US and NIES in Japan) suggesting that the prescribed subadult sampling time, culture temperature, and feeding rate will produce comparable growth measurements from future MEOGRTs. The consistency indicates that

MEOGRTs performed in different laboratories will provide similar data that will facilitate inter-test comparisons.

Concurrent with the rapid growth observed, the development of SSC occurs around 6 wpf. Therefore, if subadults are sampled at 9 weeks post fertilization (wpf), males will have high numbers of anal fin papillae (Figure 6 SSC). In addition, during this phase, gonadal development is dramatic, and there is a substantial increase in the production of vitellogenin in females. For example, the mean number of copies of *vitellogenin* RNA per ng of total liver RNA across tests increased from 2×10^5 at 6 wpf to 5×10^6 at 7 wpf. Another case illustrates this point dramatically. In the tamoxifen test, the subadults were sampled at 54 dpf in F1 and at 65 dpf in F2, a difference of 11 days. The mean liver vitellogenin protein concentration in the F1 control females was approximately 40 ng/mg liver compared with 1350 ng/mg liver in the F2 females. Because of the rapid development rate of subadult medaka, there are large quantitative changes in growth, SSC, and vitellogenin endpoints through time. Moving the subadult sampling from the 7th to the 9th wpf allows the fish the time to develop optimal mean values with low variance in all the measured endpoints.

Test temperature—The recommended temperature in the MEOGRT is 25°C, and brief excursions from the mean by individual aquaria should not be more than 2°C. Furthermore, replicates within a treatment should not statistically differ from each other, and treatments within the test should not be statistically different from each other. This recommendation is based upon the historical range of temperatures used for toxicology testing with medaka [72], and the temperature that optimized daily egg production while minimizing the occurrence of XX males. Fairly modest decreases in the test temperature reduced the fecundity of control breeding pairs. In reproductive assessments, mean fecundity was lower (22 ± 3 eggs per female per day) at 24°C than the fecundity observed at 25°C (36 ± 3 eggs per female per day). These gains in reproductive performance have large impacts on the statistical power of this endpoint simply by increasing the measureable effect size [71]. Because of this, temperatures lower than 24°C were deemed unacceptable.

Given the importance the MEOGRT places on knowing the genetic sex of individuals, the presence of genotype-phenotype mismatches, in this case XX males, has serious negative impacts on the data analysis. In fact, it has been documented in the literature [73–75], and observed in our medaka culture that a small number of unexposed XX medaka embryos will develop as normal phenotypic males, possibly caused by cortisol elevation during development [76]. At least two factors are known to increase the likelihood of these events: elevated temperature and medaka strain genetics. The published literature reports elevated rates of XX male development at 32°C, a temperature not considered for a standard protocol like the MEOGRT. Depending on strain, elevated rates are observed at 27°C, a more likely temperature for a standard testing protocol [74, 77]. Thus, we initiated an ancillary study to investigate the prevalence of XX males at temperatures suitable for medaka tests such as the MEOGRT, using the medaka strain cultured at MED (*i.e.*, orange-red). Newly fertilized eggs were cultured at various temperatures between 23°C and 32°C until hatch, then transferred to 26° water, and reared until they were sexually dimorphic. Their individual gonadal phenotype and genotype (*Dmy* status) were then assessed. Indeed, increasing temperature did increase the prevalence of XX males in the MED strain (Figure 7 XX males). Note that

higher temperatures caused an increase in the prevalence of XX males even within a range that would be appropriate for the medaka toxicity tests (*i.e.*, 24 – 26°C). The recommended temperature of 25°C for the MEOGRT balances the competing objectives of maximizing fecundity while minimizing the chances of XX males induced by temperature. In addition, because the prevalence of this phenomenon is strain dependent, efforts at MED to increase fecundity of the medaka culture by crossing separate lineages [71] have serendipitously reduced the occurrence of XX males to levels less than 1%.

Harmonized MEOGRT test guidelines have been approved by the USEPA and by the OECD [6, 7]. The final MEOGRT consists of continuously exposing parts of two generations of medaka. The F₀ generation is started on test by exposing reproductively active fish that are at least 12 wpf, for 21 days during which reproduction (*i.e.*, fecundity and fertility) is assessed. During this time, the test substance and/or its metabolites are presumed to be distributed to the gametes and tissues of the fish. After the assessment of reproduction, eggs are collected to start the F₁ generation. These fish are reared for a total of 15 wpf. A subset of F₁ fish are sampled at 9 wpf for measurement of various endpoints. Reproduction is assessed in adult F₁ fish for 21 continuous days during the period of 12 to 14 wpf. Again after the assessment of reproduction, eggs are collected and allowed to hatch, thus completing the test. The protocol specifies five test substance treatments and a dilution water control. In the F₀ generation, each test substance treatment has six replicates and the control has 12 replicates. During the first 9 weeks of the F₁ generation, this replicate structure remains the same. However, for the remainder of F₁ (once breeding pairs are established), the number of replicates per treatment is doubled to 12 replicates, and the number of replicates in the control is doubled to 24 replicates (Figure 4 MEOGRT timeline).

The MEOGRT protocol has been formulated to maximize the power of the apical endpoints of reproduction, growth and mortality while keeping implementation of the protocol feasible. The various decisions that modified the research protocol into the MEOGRT protocol were always made with the intent of providing high quality data to accurately define the dose response of these apical endpoints. At the same time, endpoints that could provide information about an AOP of a potential EDC were included in the protocol design. These non-apical endpoints are intended to characterize the AOP of the test substance at a particular level of biological organization such that the information from a MEOGRT can help identify the relevant AOP(s). For instance, decreased vitellogenin production and the development of anal fin papillae in exposed XX medaka is consistent with an androgen receptor agonist AOP. However, we believe that most important information derived from a MEOGRT dataset, for use in an ecological risk assessment, is the apical endpoints (*i.e.*, reproduction, survival, and growth).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

The research described in this document has been funded by either the US Environmental Protection Agency or the Japan Ministry of the Environment.

DISCLAIMER

The present study has been subjected to review by the National Health and Environmental Effects Research Laboratory of the US Environmental Protection Agency and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

REFERENCES

1. Manibusan MK, Touart LW. 2017 A comprehensive review of regulatory test methods for endocrine adverse health effects. *Crit Rev Toxicol* 1–49.
2. US Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention. 2011 Weight-of-evidence: Evaluating results of EDSP Tier 1 screening to identify the need for Tier 2 testing. Endocrine Disruptor Screening Program EPA-HQ-OPPT-2010–0877.
3. National Institute of Environmental Health Sciences. 1997 Validation and regulatory acceptance of toxicological test methods: A report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH publication 97–3981.
4. Organisation of Economic Co-operation and Development. 2005 Guidance document on the validation of international acceptance of new or updated test methods for hazard assessment Series on Testing and Assessment: Number 334. OECD Publishing, Paris, France.
5. US Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention. 2013 FIFRA Scientific Advisory Panel meeting: Proposed EDSP Tier 2 ecotoxicity tests. Endocrine Disruptor Screening Program EPA-HQ-OPP-2013–0182.
6. US Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention. 2015 OCSPP 890.2200: Medaka Extended One Generation Reproduction Test (MEOGRT). Endocrine Disruptor Screening Program Test Guidelines. EPA-HQ-OPPT-2014–0766-0021.
7. Organisation of Economic Co-operation and Development. 2015 Test No. 240: Medaka Extended One Generation Reproduction Test (MEOGRT). OECD Series on Testing and Assessment. OECD Publishing, Paris, France.
8. Hirshfield MF. 1980 An experimental analysis of reproductive effort and cost in the Japanese medaka, *Oryzias latipes*. *Ecology* 61:282–292.
9. Koger CS, Teh SJ, Hinton DE. 1999 Variations of light and temperature regimes and resulting effects on reproductive parameters in Medaka (*Oryzias latipes*). *Biol Reprod* 61:1287–1293. [PubMed: 10529276]
10. Urushitani H, Katsu Y, Kato Y, Tooi O, Santo N, Kawashima Y, Ohta Y, Kisaka Y, Lange A, Tyler CR, Johnson RD, Iguchi T. 2007 Medaka (*Oryzias latipes*) for use in evaluating developmental effects of endocrine active chemicals with special reference to gonadal intersex (testis-ova). *Environ Sci* 14:211–233. [PubMed: 17975535]
11. Wittbrodt J, Shima A, Scharl M. 2002 Medaka- a model organism from the far east. *Nat Rev* 3:53–64.
12. Shima A, Mitani H. 2004 Medaka as a research organism: past, present and future. *Mech Develop* 121:599–604.
13. Padilla S, Cowden J, Hinton DE, Yuen B, Law S, Kullman SW, Johnson R, Hardman RC, Flynn K, Au DWT. 2009 Use of medaka in toxicity testing. *Curr Protoc Toxicol* 39:1–36.
14. Seki M, Abe T, Maeda M, Tatarazako N, Ohnishi Y. 2003 Chapter 3: Medaka screening and definitive test for endocrine disrupting chemicals Development of Test Methods and Suitability of Medaka as Test Organism for Detection of Endocrine Disrupting Chemicals. Ministry of the Environment; Chemicals Evaluation and Research Institute, Japan.
15. Yokota H, Seki M, Maeda M, Oshima Y, Tadokoro H, Honjo T, Kobayashi K. 2001 Life-cycle toxicity of 4-nonylphenol to medaka (*Oryzias latipes*). *Environ Toxicol Chem* 20:2552–2560. [PubMed: 11699782]
16. Ankley GT, Johnson RD. 2004 Small fish models for identifying and assessing the effects of endocrine-disrupting chemicals. *ILAR J* 45:469–483. [PubMed: 15454686]

17. Koger CS, Teh SJ, Hinton DE. 2000 Determining the sensitive development stages of intersex induction in medaka (*Oryzias latipes*) exposed to 17 β -estradiol or testosterone. *Mar Environ Res* 50:201–206. [PubMed: 11460690]
18. Orn S, Yamani S, Norrgren L. 2006 Comparison of vitellogenin induction, sex ratio, and gonad morphology between zebrafish and Japanese medaka after exposure to 17 α -ethinylestradiol and 17 β -trenbolone. *Arch Environ Contam Toxicol* 51:237–243. [PubMed: 16783621]
19. Ankley GT, Jensen KM, Makynen EA, Kahl MD, Korte JJ, Hornung MW, Henry TR, Denny JS, Leino RL, Wilson VS, Cardon MC, Hartig PC, Gray LE. 2003 Effects of the androgenic growth promoter 17-beta-trenbolone on fecundity and reproductive endocrinology of the fathead minnow. *Environ Toxicol Chem* 22:1350–1360. [PubMed: 12785594]
20. Seki M, Yokots H, Matsubara H, Maeda M, Tadokoro H, Kobayashi K. 2004 Fish full life-cycle testing for androgen methyltestosterone on medaka (*Oryzias latipes*). *Environ Toxicol Chem* 23:774–781. [PubMed: 15285372]
21. Haselman JT, Kosian PA, Korte JJ, Olmstead AW, Iguchi T, Johnson RD, Degitz SJ. 2015 Development of the larval amphibian growth and development assay: effects of chronic 4-*tert*-octylphenol or 17 β -trenbolone exposure in *Xenopus laevis* from embryo to juvenile. *J Appl Toxicol*.
22. Hara Y, Strussmann CA, Hashimoto S. 2007 Assessment of short-term exposure to nonylphenol in Japanese medaka using sperm velocity and frequency of motile sperm. *Arch Environ Contam Toxicol* 53:406–410. [PubMed: 17657460]
23. Crofton KM, Paul KB, DeVito MJ, Hedge JM. 2007 Short-term in vivo exposure to the water contaminant triclosan: evidence for disruption of thyroxine. *Environ Toxicol Pharmacol* 24:194–197. [PubMed: 21783810]
24. Denslow N, Sepúlveda M. 2007 Ecotoxicological effects of endocrine disrupting compounds on fish reproduction In Babin PJ, Cerda J, Lubzens E, eds, *The Fish Oocyte: From Basic Studies to Biotechnological Applications*. Springer, Netherlands, pp 255–322.
25. Seki M, Yokota H, Matsubara H, Maeda M, Tadokoro H, Kobayashi K. 2003 Fish full life-cycle testing for the weak estrogen 4-*tert*-pentylphenol on medaka (*Oryzias latipes*). *Environ Toxicol Chem* 22:1487–1496. [PubMed: 12836973]
26. McCarty LS, Mackay D, Smith AD, Ozburn GW, Dixon DG. 1992 Residue-based interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: neutral narcotic organics. *Environ Toxicol Chem* 11:917–930.
27. Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N, Hori H, Hamaguchi S, Sakaizumi M. 2002 *DMY* is a y-specific dm-domain gene required for male development in the medaka fish. *Nature* 417:559–563. [PubMed: 12037570]
28. Flynn K, Swintek J, Johnson R. 2013 Use of gene expression data to determine effects on gonad phenotype in Japanese medaka after exposure to trenbolone or estradiol. *Environ Toxicol Chem* 32:1344–1353. [PubMed: 23423942]
29. Tatarazako N, Koshio M, Hori H, Morita M, Iguchi T. 2004 Validation of an enzyme-linked immunosorbent assay method for vitellogenin in the medaka. *J Health Sci* 50:301–308.
30. Organisation of Economic Co-operation and Development. 2006 Report of the validation of the 21 day fish screening assay for the detection of endocrine substances (phase 1b) In *OECD Series on Testing and Assessment: Number 61*. OECD Publishing, Paris, France.
31. Johnson R, Wolf J, Braunbeck T. 2009 Organisation of Economic Co-operation and Development, Guidance Document for the Diagnosis of Endocrine-Related Histopathology of Fish Gonads. OECD Publishing, Paris, France.
32. Green JW, Springer TA, Saulnier AN, Swintek J. 2014 Statistical analysis of histopathological endpoints. *Environ Toxicol Chem* 33:1108–1116. [PubMed: 24464649]
33. Organisation of Economic Co-operation and Development. 2006 Current approaches in the statistical analysis of ecotoxicity data: A guidance to application *OECD Series on Testing and Assessment: Number 54*, OECD Publishing, Paris, France.
34. Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrano JA, Tietge JE, Villeneuve DL. 2010 Adverse outcome

- pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* 29:730–741. [PubMed: 20821501]
35. Cripe GM, Hemmer BL, Goodman LR, Fournie JW, Raimondo S, Vennari JC, Danner RL, Smith K, Manfredonia BR, Kulaw DH, Hemmer MJ. 2009 Multigenerational exposure of the estuarine sheepshead minnow (*Cyprinodon variegatus*) to 17 β -estradiol. I. Organism-level effects over three generations. *Environ Toxicol Chem* 28:2397–2408. [PubMed: 19839653]
 36. Hirai N, Nanba A, Koshio M, Kondo T, Morita M, Tatarazako N. 2006 Feminization of Japanese medaka (*Oryzias latipes*) exposed to 17 β -estradiol: formation of testis-ova and sex transformation during early ontogeny. *Aquat Toxicol* 77:78–86. [PubMed: 16330109]
 37. Raimondo S, Hemmer BL, Goodman LR, Cripe GM. 2009 Multigenerational exposure of the estuarine sheepshead minnow (*Cyprinodon variegatus*) to 17 β -estradiol. II. Population-level effects through two life cycles. *Environ Toxicol Chem* 28:2409–2415. [PubMed: 19586131]
 38. Seki M, Fujishima S, Nozaka T, Maeda M, Kobayashi K. 2006 Comparison of response to 17 β -estradiol and 17 β -trenbolone among three small fish species. *Environ Toxicol Chem* 25:2742–2752. [PubMed: 17022416]
 39. Seki M, Yokota H, Maeda M, Kobayashi K. 2005 Fish full life-cycle testing for 17 β -estradiol on medaka. *Environ Toxicol Chem* 24:1259–1266. [PubMed: 16111009]
 40. Seki M, Yokota H, Maeda M, Tadokoro H, Kobayashi K. 2003 Effects of 4-nonylphenol and 4-*tert*-octylphenol on sex differentiation and vitellogenin induction in medaka (*Oryzias latipes*). *Environ Toxicol Chem* 22:1507–1516. [PubMed: 12836975]
 41. Gray MA, Metcalfe CD. 1999 Toxicity of 4-*tert*-octylphenol to early life stages of Japanese medaka (*Oryzias latipes*). *Aquat Toxicol* 46:149–154.
 42. Gronen S, Denslow N, Manning S, Barnes S, Barnes D, Brouwer M. 1999 Serum vitellogenin levels and reproductive impairment of male Japanese Medaka (*Oryzias latipes*) exposed to 4-*tert*-octylphenol. *Environ Health Perspect* 107:385–390. [PubMed: 10210694]
 43. Korner W, Spengler P, Bolz U, Schuller W, Hanf V, Metzger JW. 2001 Substances with estrogenic activity in effluents of sewage treatment plants in southwestern Germany. 2. Biological analysis. *Environ Toxicol Chem* 20:2142–2151. [PubMed: 11596742]
 44. Laws SC, Carey SA, Ferrell JM, Bodman GJ, Cooper RL. 2000 Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats. *Toxicol Sci* 54:154–167. [PubMed: 10746942]
 45. Gray MA, Nimi AJ, Metcalfe CD. 1999 Factors affecting the development of testis-ova in medaka, *Oryzias latipes*, exposed to octylphenol. *Environ Toxicol Chem* 18:1835–1842.
 46. Knorr S, Braunbeck T. 2002 Decline in reproductive success, sex reversal, and developmental alterations in Japanese medaka (*Oryzias latipes*) after continuous exposure to octylphenol. *Ecotoxicol Environ Saf* 51:187–196. [PubMed: 11971639]
 47. Metcalfe TL, Metcalfe CD, Kiparissis Y, Nimi AJ, Foran CM, Benson WH. 2000 Gonadal development and endocrine responses in Japanese medaka (*Oryzias Latipes*) exposed to *o,p'*-DDT in water or through maternal transfer. *Environ Toxicol Chem* 19:1893–1900.
 48. Uchida M, Nakamura H, Kagami Y, Kusano T, Arizono K. 2010 Estrogenic effects of *o,p'*-DDT exposure in Japanese medaka (*Oryzias latipes*). *J Toxicol Sci* 35:605–608. [PubMed: 20686349]
 49. Schmieder PK, Tapper MA, Denny JS, Kolanczyk RC, Sheedy BR, Henry TR, Veith GD. 2004 Use of trout liver slices to enhance mechanistic interpretation of estrogen receptor binding for cost-effective prioritization of chemicals within large inventories. *Environ Sci Technol* 38:6333–6342. [PubMed: 15597890]
 50. Nishihara T, Nishikawa J, Kanayama T, Dakeyama F, Saito K, Imagawa M, Takatori S, Kitagawa Y, Hori S, Utsumi H. 2000 Estrogenic activities of 517 chemicals by yeast two-hybrid assay. *J Health Sci* 46:282–298.
 51. Shang Y, Brown M. 2002 Molecular determinants for the tissue specificity of SERMs. *Science* 295:2465–2468. [PubMed: 11923541]
 52. Chikae M, Ikeda R, Hasan Q, Morita Y, Tamiya E. 2004 Effects of tamoxifen, 17 α -ethynylestradiol, flutamide, and methyltestosterone on plasma vitellogenin levels of male and female Japanese medaka (*Oryzias latipes*). *Environ Toxicol Pharmacol* 17:29–33. [PubMed: 21782710]

53. Sun L, Zha J, Spear PA, Wang Z. 2007 Tamoxifen effects on the early life stages and reproduction of Japanese medaka (*Oryzias latipes*). *Environ Toxicol Pharmacol* 24:23–29. [PubMed: 21783785]
54. Singh R, Singh AK, Tripathi M. 2012 Effect of a non-steroidal tamoxifen on the gonad and sex differentiation in Nile tilapia, *Oreochromis niloticus*. *J Environ Biol* 33:799–803. [PubMed: 23360010]
55. Sone K, Hinago M, Itamoto M, Katsu Y, Watanabe H, Urushitani H, Tooi O, Guillette LJ Jr, Iguchi T. 2005 Effects of an androgenic growth promoter 17 β -trenbolone on masculinization of mosquitofish (*Gambusia affinis affinis*). *Gen Comp Endocrinol* 143:151–160. [PubMed: 16061073]
56. Wilson VS, Lambright C, Ostby J, Gray LE Jr. 2002 *In Vitro* and *in Vivo* effects of 17 β -trenbolone: A feedlot effluent contaminant. *Toxicol Sci* 70:202–211. [PubMed: 12441365]
57. Cripe GM, Hemmer BL, Raimondo S, Goodman LR, Kulaw DH. 2010 Exposure of three generations of the estuarine sheepshead minnow (*Cyprinodon variegatus*) to the androgen, 17 β -trenbolone: Effects on survival, development, and reproduction. *Environ Toxicol Chem* 29:2079–2087. [PubMed: 20821666]
58. Ogino Y, Hirakawa I, Inohaya K, Sumiya E, Miyagawa S, Denslow N, Yamada G, Tatarazako N, Iguchi T. 2014 *Bmp7* and *Lef1* are the downstream effectors of androgen signaling in androgen-induced sex characteristics development in medaka. *Endocrinology* 155:449–462. [PubMed: 24248458]
59. Bayley M, Larsen PF, Baekgaard H, Baatrup E. 2003 The effects of vinclozolin, an anti-androgenic fungicide, on male guppy secondary sex characters and reproductive success. *Biol Reprod* 69:1951–1956. [PubMed: 12930724]
60. Makynen EA, Kahl MD, Jensen KM, Tietge JE, Wells KL, Van Der Kraak G, Ankley GT. 2000 Effects of the mammalian anti-androgen vinclozolin on development and reproduction of the fathead minnow (*Pimephales promelas*). *Aquat Toxicol* 48:461–475. [PubMed: 10794831]
61. Martinovic D, Blake LS, Durhan EJ, Greene KJ, Kahl MD, Jensen KM, Makynen EA, Villeneuve DL, Ankley GT. 2008 Reproductive toxicity of vinclozolin in the fathead minnow: confirming an anti-androgenic mode of action. *Environ Toxicol Chem* 27:478–488. [PubMed: 18348629]
62. Sebillot A, Damdimopoulou P, Ogino Y, Spirhanzlova P, Miyagawa S, Du Pasquier D, Moutassim N, Iguchi T, Lemkine GF, Demeneix BA, Tindall AJ. 2014 Rapid fluorescent detection of (anti)androgens with *spiggin-gfp* medaka. *Environ Sci Technol* 48:10919–10928. [PubMed: 25171099]
63. Nakamura A, Takanobu H, Tamura I, Yamamuro M, Iguchi T, Tatarazako N. 2014 Verification of responses of Japanese medaka (*Oryzias latipes*) to anti-androgens, vinclozolin and flutamide, in short-term assays. *J Appl Toxicol* 34:545–553. [PubMed: 24072643]
64. Kelce WR, Monosson E, Gamcsik MP, Laws SC, Gray LE. 1994 Environmental hormone disruptors: evidence that vinclozolin developmental toxicity is mediated by anti-androgenic metabolites. *Toxicol Appl Pharmacol* 126:276–285. [PubMed: 8209380]
65. Wong C, Kelce WR, Sar M, Wilson EM. 1995 Androgen receptor antagonist versus agonist activities of the fungicide vinclozolin relative to hydroxyflutamide. *J Biol Chem* 270:19998–20003. [PubMed: 7650017]
66. Hinfray N, Palluel O, Turies C, Cousin C, Porcher JM, Brion F. 2006 Brain and gonadal aromatase as potential targets of endocrine disrupting chemicals in a model species, the zebrafish (*Danio rerio*). *Environ Toxicol* 21:332–337. [PubMed: 16841311]
67. Monod G, De Mones A, Fostier A. 1993 Inhibition of ovarian microsomal aromatase and follicular oestradiol secretion by imidazole fungicides in rainbow trout. *Mar Environ Res* 35:153–157.
68. Skolness SY, Durhan E, Garcia-Reyero N, Jensen KM, Kahl MD, Makynen EA, Martinovic-Weigelt D, Perkins E, Villeneuve DL, Ankley GT. 2011 Effects of a short-term exposure to the fungicide, prochloraz, on endocrine function and gene expression in female fathead minnows (*Pimephales promelas*). *Aquat Toxicol* 103:170–178. [PubMed: 21470553]
69. Ankley GT, Jensen KM, Durhan EJ, Makynen EA, Butterworth BC, Kahl MD, Villeneuve DL, Linnam A, Gray LE, Cardon M, Wilson VS. 2005 Effects of two fungicides with multiple modes

- of action on reproductive endocrine function in the fathead minnow (*Pimephales promelas*). *Toxicol Sci* 86:300–308. [PubMed: 15901916]
70. Schiller V, Zhang X, Hecker M, Schafers C, Fischer R, Fenske M. 2014 Species-specific considerations in using the fish embryo test as an alternative to identify endocrine disruption. *Aquat Toxicol* 155C:62–72.
71. Flynn K, Swintek J, Johnson R. 2017 The influence of control group reproduction on the statistical power of the Environmental Protection Agency's Medaka Extended One Generation Reproduction Test (MEOGRT). *Ecotoxicol Environ Saf* 136:8–13. [PubMed: 27810580]
72. Denny JS, Spehar RL, Mead KE, Yousuff SC. 1991 Guidelines for culturing the Japanese medaka, *Oryzias latipes*. USEPA/600/3–91/064.
73. Nanda I, Hornung U, Kondo M, Schmid M, Schartl M. 2003 Common spontaneous sex-reversed XX males of the medaka *Oryzias latipes*. *Genetics* 163:245–251. [PubMed: 12586712]
74. Sato T, Endo T, Yamahira K, Hamaguchi S, Sakaizumi M. 2005 Induction of female-to-male sex reversal by high temperature treatment in medaka, *Oryzias latipes*. *Zoolog Sci* 22:985–988. [PubMed: 16219978]
75. Selim KM, Shinomiya A, Otake H, Hamaguchi S, Sakaizumi M. 2009 Effects of high temperature on sex differentiation and germ cell population in medaka, *Oryzias latipes*. *Aquaculture* 289:340–349.
76. Hayashi Y, Kobira H, Yamaguchi T, Shiraishi E, Yazawa T, Hirai T, Kamei Y, Kitano T. 2010 High temperature causes masculinization of genetically female medaka by elevation of cortisol. *Mol Reprod Dev* 77:679–686. [PubMed: 20653000]
77. Hattori RS, Gould RJ, Fujioka T, Saito T, Kurita J, Strussmann CA, Yokota M, Watanabe S. 2007 Temperature-dependent sex determination in Hd-rR medaka *Oryzias latipes*: gender sensitivity, thermal threshold, critical period, and *DMRT1* expression profile. *Sex Dev* 1:138–146. [PubMed: 18391524]

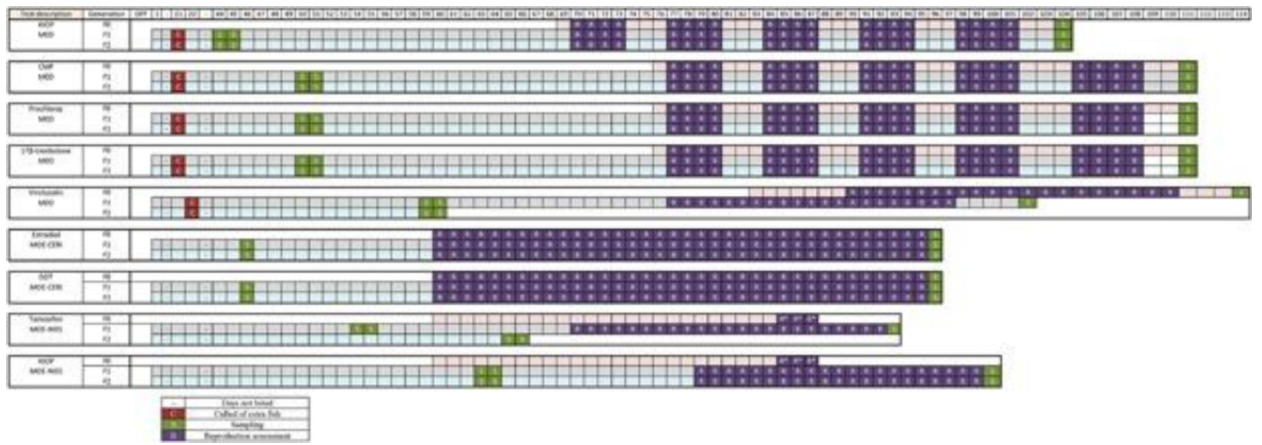


Figure 1. The timelines for each of the presented tests showing the duration, as measured in days post fertilization (dpf), of each generation and the key events (i.e., culling, sampling, and reproduction assessment) within each generation. The tested chemical and the laboratory that completed the test are listed.

EPA Author Manuscript

EPA Author Manuscript

EPA Author Manuscript

Research protocol exposure and endpoint timeline																															
Generation																															
F ₀	1	2	3	4	5																										
F ₁			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15														
F ₂																1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Study week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29		
Endpoints																															
Fecundity	F ₀					F ₁										F ₂															
Fertility	F ₀					F ₁										F ₂															
Hatching				F ₁												F ₂															
Survival				F ₀	F ₁				F ₁								F ₁	F ₂					F ₂								F ₂
Growth				F ₀				F ₁									F ₁					F ₂								F ₂	
Vitellogenin				F ₀				F ₁									F ₁					F ₂								F ₂	
Secondary sex				F ₀				F ₁									F ₁					F ₂								F ₂	
Histopatholog				F ₀				F ₁									F ₁					F ₂								F ₂	
Study week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29		

Life-stage key

Embryo

Eleutheroembryo

Juvenile

Subadult

Adult

6 groups of replicates

- 5 for test substance treatments
- 1 for controls

Within-group design

- 6 replicate aquaria

Figure 2. The timeline and endpoints of the described research protocol that was the starting point for developing the MEOGRT. The duration of the in-life exposure is 29 wk and spans 2 full generations. The F₀ generation serves as a loading phase for the F₁ embryos. The test is terminated after reproduction in the F₂ generation. The rows to the right of each endpoint indicate the study week in which the endpoint is measured in each generation.

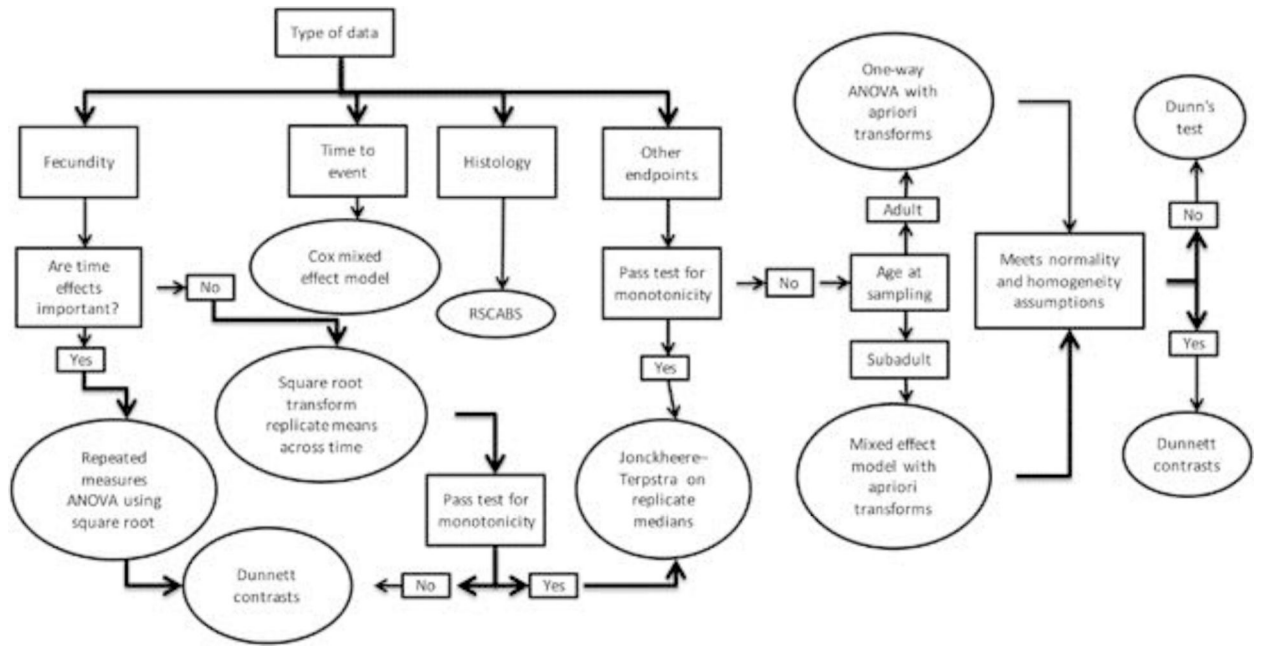


Figure 3. A flow chart for the recommended statistical procedures for MEOGRT data analysis. ANOVA = analysis of variance; RSCABS = Rao–Scott Cochran–Armitage by Slices.

MEOGRT exposure and endpoint timeline																					
Generation																					
F ₀	1	2	3	4															Life-stage key Embryo Eleutheroembryo Juvenile Subadult Adult 7 groups of replicates • 5 for test substance treatments • 2 for controls (4 if solvent is used) Within-group design • 12 reps for reproduction, pathology, growth, SSC (wks 10-18) • 6 reps for hatching, survival, vitellogenin, growth, (wks 1-9)		
F ₁				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
F ₂																	1	2			
Study week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19	
	Endpoints																				
Fecundity	F ₀														F ₁						
Fertility	F ₀														F ₁						
Hatching					F ₁																F ₂
Survival																					F ₁
Growth					F ₀																F ₁
Vitellogenin																					F ₁
Secondary sex																					F ₁
Histopathology																					F ₁
Study week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19	

Figure 4. The timeline and endpoint summary for the final MEOGRT protocol. The duration of the in-life exposure is 19 wk and spans 1 full generation (F₁). The F₀ generation serves as a loading phase for the F₁ embryos. The test is terminated after the F₂ embryos hatch. The rows to the right of each endpoint indicate the study week in which the endpoint is measured in each generation. SSC = secondary sex character.

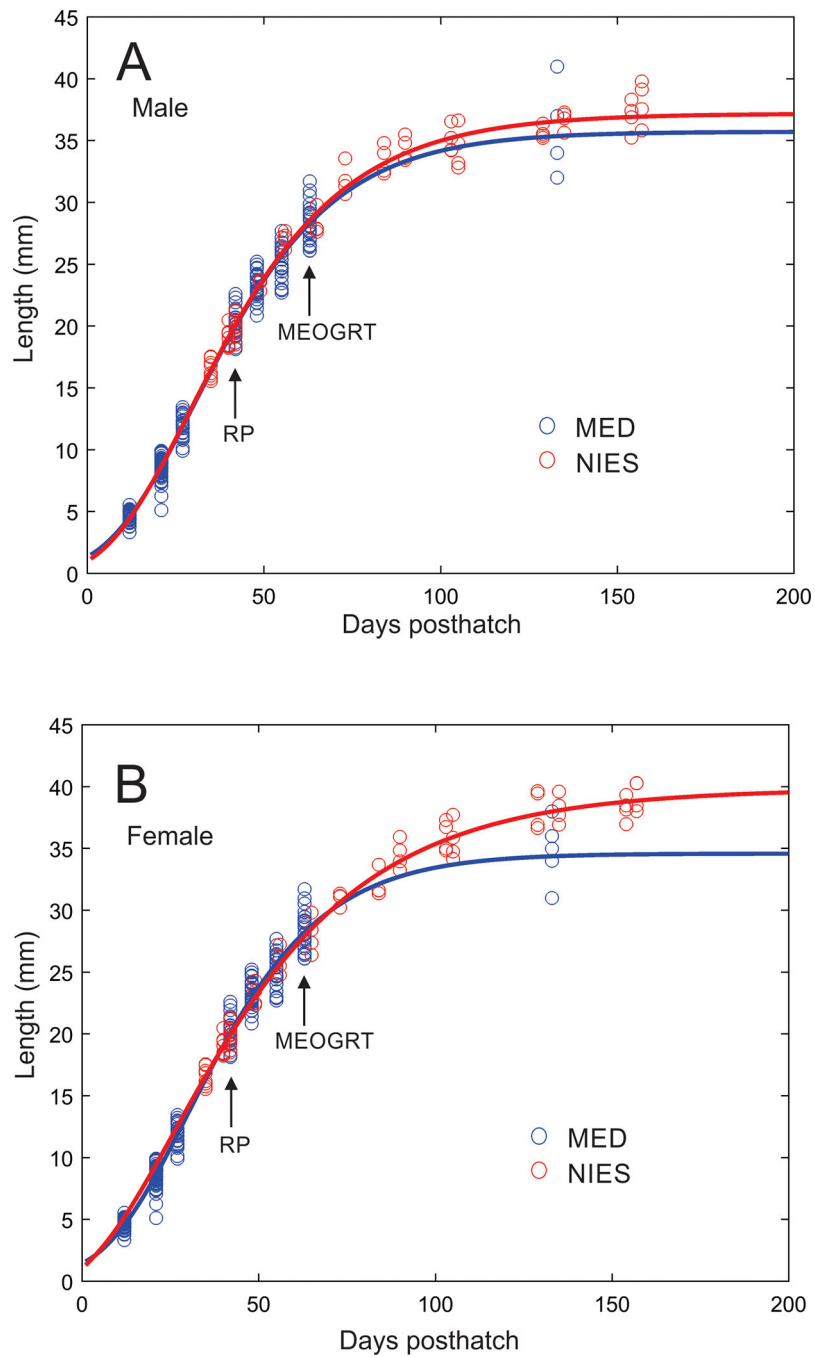


Figure 5. Growth, as measured by total length, from hatch to adult of female (A) and male (B) medaka reared at the USEPA Mid-Continent Ecology Division (MED) or Japan's National Institute for Environmental Studies (NIES) under the conditions specified in the MEOGRT protocol. Circles are measured data points, and solid curves are modeled growth based on the data points. Arrows indicate the subadult sampling time with either the research protocol (RP) or the final MEOGRT protocol (MEOGRT).

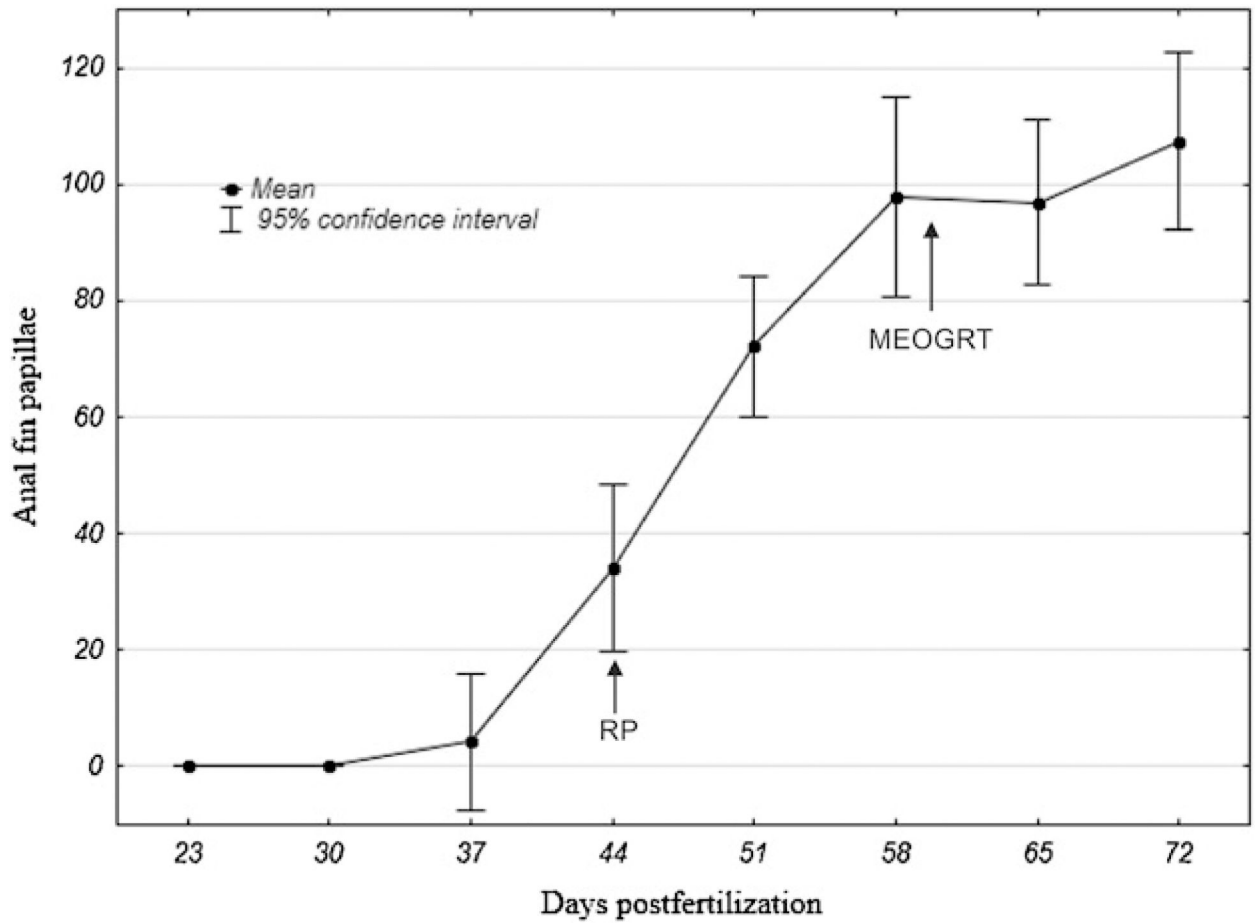


Figure 6. Secondary sexual characteristic (SSC) development in males through time when male medaka are reared under the conditions specified in the MEOGRT protocol. The numbers of anal fin papillae, the measurement of secondary sex character, were counted per male. Arrows indicate the subadult sampling time with either the research protocol (RP) or the final MEOGRT protocol (MEOGRT).

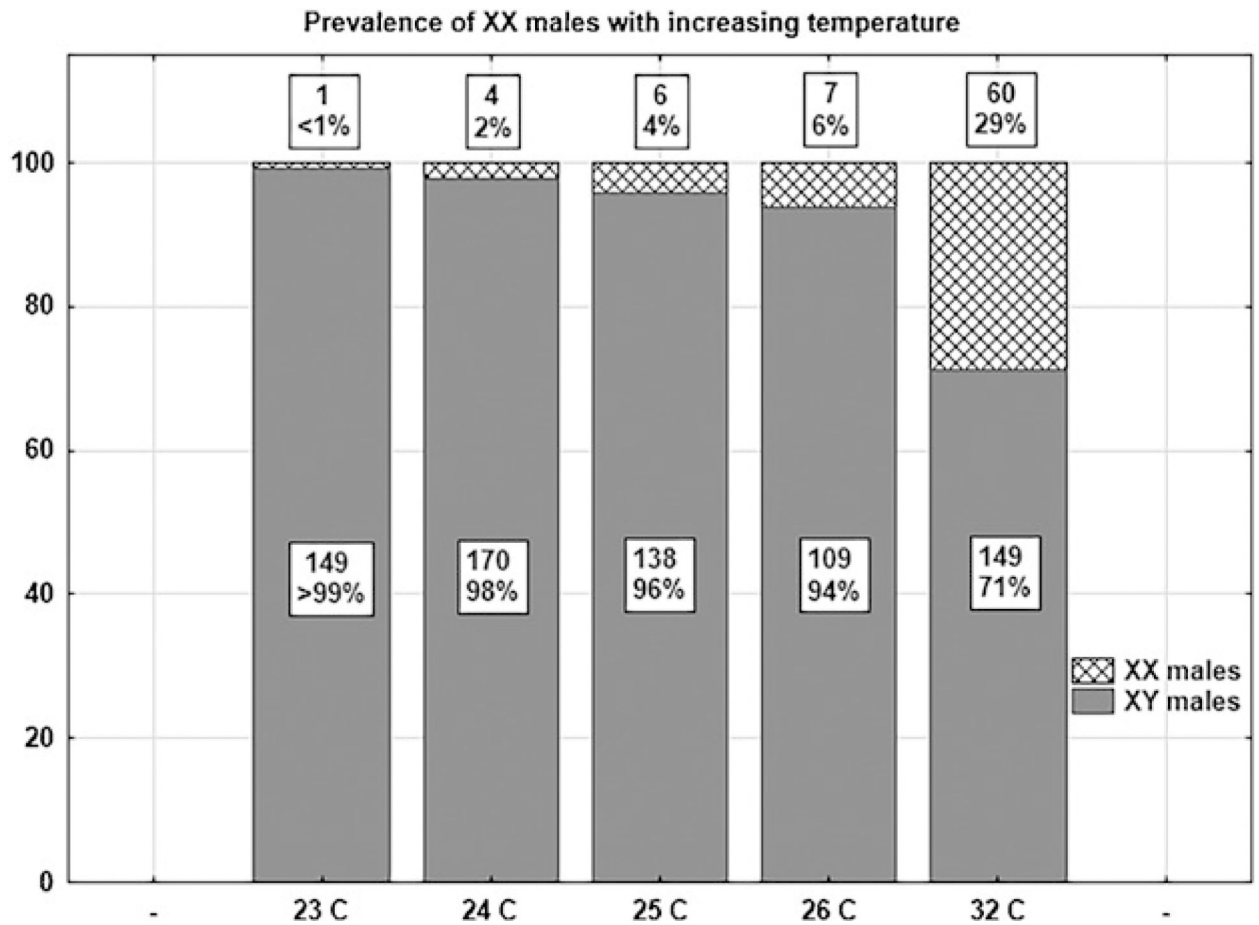


Figure 7.

Prevalence, as a percentage, of XX males with increasing temperature during development. The bars show the proportion of phenotypic males that are either genotypic XX or XY. The box above each bar contains number and the percentage of XX males. The box within each bar lists the number of the XY males.

Table 1.

Summary information for the chemicals used in the nine presented medaka multi-generational tests.

Chemical	Mode of Action	Quantification Method	Test Concentrations (mean \pm standard deviation)				
17 β -estradiol	ER agonist	SPE,RIA	0.92 \pm 0.11 ng/L	2.8 \pm 0.35 ng/L	8.9 \pm 1.0 ng/L	27.9 \pm 2.9 ng/L	84.3 \pm 7.1 ng/L
4- <i>t</i> -octylphenol (MED)	ER agonist	LC or GC	6.2 \pm 0.6 μ g/L	12.5 \pm 1.5 μ g/L	25.1 \pm 2.6 μ g/L	51.1 \pm 5.6 μ g/L	102.3 \pm 8.5 μ g/L
4- <i>t</i> -octylphenol (NIES)	ER agonist	none	6.25 μ g/L	12.5 μ g/L	25 μ g/L	50 μ g/L	100 μ g/L
<i>o,p'</i> -DDT	ER agonist	GC-MS	0.03 \pm 0.004 μ g/L	0.07 \pm 0.01 μ g/L	0.22 \pm 0.04 μ g/L	0.69 \pm 0.17 μ g/L	1.90 \pm 0.43 μ g/L
4-chloro-3-methylphenol	ER agonist	HPLC	20.7 \pm 4.6 μ g/L	43.3 \pm 14.1 μ g/L	88.0 \pm 27.7 μ g/L	165.1 \pm 22.1 μ g/L	344.7 \pm 74.4 μ g/L
tamoxifen	ER antagonist	GC-MS	1.32 \pm 0.19 μ g/L	2.51 \pm 0.28 μ g/L	5.08 \pm 0.65 μ g/L	10.17 \pm 1.34 μ g/L	20.44 \pm 2.58 μ g/L
17 β -trenbolone	AR agonist	SPE, HPLC	2.2 \pm 0.5 ng/L	5.1 \pm 1.2 ng/L	12.9 \pm 2.7 ng/L	31.7 \pm 4.8 ng/L	84.2 \pm 13.0 ng/L
vinclozolin	AR antagonist	HPLC	17.3 \pm 1.6 μ g/L	33.2 \pm 3.4 μ g/L	69.5 \pm 6.2 μ g/L	136.3 \pm 17.6 μ g/L	253.3 \pm 58.7 μ g/L
prochloraz	Steroidogenesis inhibitor	HPLC	5.3 \pm 1.1 μ g/L	9.2 \pm 1.7 μ g/L	17.5 \pm 3.1 μ g/L	25.0 \pm 3.8 μ g/L	41.1 \pm 3.4 μ g/L

ER = estrogen receptor; AR = androgen receptor; SPE = solid phase extraction; RIA = radioimmunoassay; LC = liquid chromatography; GC = gas chromatography; MS = mass spectrometry; HPLC = high-performance liquid chromatography

Table 2.

The LOECs for the measured endpoints from the medaka multi-generational tests on chemicals that are estrogen receptor agonists.

Endpoint	Estrogen Receptor Agonist											
	E2 CER1		4HOP EPA		4HOP NIES		<i>o,p'</i> -DDT CER1		CMP EPA			
	% Change	LOEC	% Change	LOEC	% Change	LOEC	% Change	LOEC	% Change	LOEC		
Fecundity	-30%	28 ng/L	-90%	102 µg/L	-80%	100 µg/L		> 1.9 µg/L		> 345 µg/L		
Fertility	-60%	28 ng/L	-70%	51 µg/L	-99%	100 µg/L	-50%	1.9 µg/L		> 345 µg/L		
Hatch		> 84 ng/L	-50%	102 µg/L	-20%	25 µg/L		> 1.9 µg/L		> 345 µg/L		
VTG	410000	28 ng/L	16800%	25 µg/L	400%	6.25 µg/L	200%	0.22 µg/L	40%	345 µg/L		
SSC	-94%	28 ng/L	-50%	6 µg/L	-90%	6.25 µg/L	-40%	1.9 µg/L		> 345 µg/L		
Weight		> 84 ng/L	20%	51 µg/L	20%	25 µg/L		> 1.9 µg/L		> 345 µg/L		
Length		> 84 ng/L	10%	13 µg/L		> 100 µg/L		> 1.9 µg/L	-10%	165 µg/L		
Larval survival	NA	NA	NA	NA	NA	NA		> 1.9 µg/L		> 345 µg/L		
Subadult survival		> 84 ng/L		> 102 µg/L		> 100 µg/L		> 1.9 µg/L		> 345 µg/L		
Adult survival		> 84 ng/L		> 102 µg/L	-70%	100 µg/L		> 1.9 µg/L		> 345 µg/L		

Each LOEC is color-coded with red representing a statistically significant decrease in the measurement as compared to controls and green representing a statistically significant increase in the measurement as compared to controls. No color is used when a LOEC was not observed. The percent change is also listed to provide information on the magnitude of the observed response.

Table 3.

The LOEC for each of the measured endpoints for those medaka multi-generational tests that used chemicals that were not estrogen receptor agonists (*i.e.*, ER antagonist, AR agonist, AR antagonist, or steroid inhibitor).

Endpoint	ER Antagonist		AR Agonist		AR Antagonist		Steroid Inhibitor	
	Tamoxifen NIES		17 β -trenbolone EPA		Vinclozolin EPA		Prochloraz EPA	
	% Change	LOEC	% Change	LOEC	% Change	LOEC	% Change	LOEC
Fecundity	-60%	20 μ g/L	-100%	32 ng/L	-98%	253 μ g/L	-70%	25 μ g/L
Fertility	-10%	10 μ g/L	-50%	13 ng/L	-100%	253 μ g/L		> 41 μ g/L
Hatch	-20%	2.5 μ g/L		> 84 ng/L		> 253 μ g/L	-30%	25 μ g/L
VTG	2000%	1.3 μ g/L	-60%	32 ng/L	2300%	136 μ g/L	-96%	9 μ g/L
SSC	-30%	10 μ g/L	0 to 39	32 ng/L	-49%	33 μ g/L	-30%	9 μ g/L
Weight	-30%	20 μ g/L		> 84 ng/L	20%	33 μ g/L	-30%	9 μ g/L
Length	-20%	20 μ g/L		> 84 ng/L	10%	70 μ g/L	-10%	9 μ g/L
Larval survival		NA	NA	NA	-20%	33 μ g/L	-60%	25 μ g/L
Subadult survival		NA		> 84 ng/L		> 253 μ g/L		> 25 μ g/L
Adult survival		>20 μ g/L		> 84 ng/L		> 253 μ g/L		> 41 μ g/L

Each LOEC is color-coded with red representing a statistically significant decrease in the measurement as compared to controls and green representing a statistically significant increase in the measurement as compared to controls. No color is used when a LOEC was not observed. The percent change is also listed to provide information on the magnitude of the observed response.

Table 4.

The statistical power of three test designs showing the increased power provided by the final MEOGRT protocol at various reductions in fecundity.

Fecundity Reduction (%)	Research Protocol	Considered Protocol	Published MEOGRT
	Balanced Design Control Reps: 6 Treatment Reps: 6	2:1 Design Control Reps: 12 Treatment Reps: 6	2:1 Design Control Reps: 24 Treatment Reps: 12
20	0.17	0.15	0.40
30	0.38	0.44	0.83
40	0.66	0.76	0.99
50	0.83	0.92	1
70	0.91	0.96	1