Identification of Putative Nuclear Receptors and Steroidogenic Enzymes in Murray-Darling Rainbowfish (*Melanotaenia fluviatilis*) Using RNA-Seq and *De Novo* Transcriptome Assembly

Peter A. Bain[1]*, Alexie Papanicolaou[2¤], Anupama Kumar[1]

1 Commonwealth Scientific and Industrial Research Organisation (CSIRO), Division of Land and Water, Urrbrae, South Australia, Australia, 2 Commonwealth Scientific and Industrial Research Organisation (CSIRO), Division of Land and Water, Black Mountain, Australian Capital Territory, Australia

¤ Current address: Hawkesbury Institute for the Environment, Western Sydney University, Richmond, New South Wales, Australia
* peter.bain@csiro.au

Abstract

Murray-Darling rainbowfish (*Melanotaenia fluviatilis* [Castelnau, 1878]; Atheriniformes: Melanotaeniidae) is a small-bodied teleost currently under development in Australasia as a test species for aquatic toxicological studies. To date, efforts towards the development of molecular biomarkers of contaminant exposure have been hindered by the lack of available sequence data. To address this, we sequenced messenger RNA from brain, liver and gonads of mature male and female fish and generated a high-quality draft transcriptome using a *de novo* assembly approach. 149,742 clusters of putative transcripts were obtained, encompassing 43,841 non-redundant protein-coding regions. Deduced amino acid sequences were annotated by functional inference based on similarity with sequences from manually curated protein sequence databases. The draft assembly contained protein-coding regions homologous to 95.7% of the complete cohort of predicted proteins from the taxonomically related species, *Oryzias latipes* (Japanese medaka). The mean length of rainbowfish protein-coding sequences relative to their medaka homologues was 92.1%, indicating that despite the limited number of tissues sampled a large proportion of the total expected number of protein-coding genes was captured in the study. Because of our interest in the effects of environmental contaminants on endocrine pathways, we manually curated subsets of coding regions for putative nuclear receptors and steroidogenic enzymes in the rainbowfish transcriptome, revealing 61 candidate nuclear receptors encompassing all known subfamilies, and 41 putative steroidogenic enzymes representing all major steroidogenic enzymes occurring in teleosts. The transcriptome presented here will be a valuable resource for researchers interested in biomarker development, protein structure and function, and contaminant-response genomics in Murray-Darling rainbowfish.

Introduction

The Murray-Darling rainbowfish (*Melanotaenia fluviatilis* [Castelnau, 1878]; Atheriniformes: Melanotaeniidae) is a small-bodied freshwater fish endemic to the Murray-Darling basin, a region that produces approximately half of the Australia's total irrigated agricultural output [1]. Due to its small body size, *M. fluviatilis* is readily maintained in laboratory aquaria and has been used to study the aquatic toxicology of organic and inorganic contaminants including pesticides and herbicides [2,3,4,5,6,7], heavy metals [8], and crude oil [9,10], and has been deployed in a field-based mobile laboratory to study effluent toxicity on-site [11]. Prominent secondary sexual characteristics combined with a relatively short maturation time make *M. fluviatilis* well-suited to investigating the sex-specific effects of endocrine disrupting chemicals (EDCs) on reproduction and sex-specific biomarkers [12,13,14,15,16,17,18,19].

The past two decades has seen increasing interest in the potential adverse effects of EDCs in aquatic organisms. The major focus to date has been on estrogenic EDCs, which have been shown to induce the expression of female-specific biomarkers in male fish and ultimately result in intersex condition in some species [20,21,22,23,24]. Small-bodied fish species have gained popularity as model organisms for EDC research in Asia (zebrafish and Japanese medaka), Europe (roach), and North America (fathead minnow). Through the use of estrogen-responsive biomarkers, most notably hepatic vitellogenin expression, variable inter-species sensitivity to natural and synthetic estrogens has been observed in laboratory-based tests [25]. In the case of xenoestrogens, inter-species differences sensitivity can be predicted using *in vitro* reporter assays [25] and are least partly explained by sequence diversity in the ligand binding domain of estrogen receptor α (ERα) [26], suggesting that characterising the phylogenetics and *in vitro* sensitivity of key receptors can provide insight into the potential susceptibility of a given fish species to the effects of EDCs. Since the environmental persistence of endocrine-active contaminants such as pharmaceuticals varies with temperature, sunlight and water chemistry [27] estimation of potential risks is likely to be improved through the use of endemic species as test organisms and by taking local environmental conditions into consideration during experimental design. While some progress has been made with regard to the effects of EDCs on endemic species in the southern hemisphere, a recent report highlighted the need for further research in native freshwater fish species in Australia [28].

A teleost-specific genome duplication (TSGD) event [29,30,31,32] is thought to have contributed to the extreme physiological and genetic diversity present in extant members of this lineage, which is estimated to comprise more than 26,000 species [33]. The maintenance and subsequent sequence divergence of gene pairs arising from the TSGD has in many cases resulted in the functional differentiation of paralogous genes. For example, two copies of the *ESR2* gene, encoding ERβ, have been maintained in most teleosts, with the two receptor subtypes ERβ1 and ERβ2 responding with different sensitivity to estrogens and xenoestrogens [34,35]. Many acanthomorpha (spiny-finned fishes including medaka and rainbowfish) are also known to express two androgen receptor (AR) subtypes, designated ARα and ARβ, while in the otomorpha group (which includes cyprinids such as zebrafish and roach) a single AR subtype similar to ARβ is present [36]. We have recently shown that ARα and ARβ from Murray-Darling rainbowfish exhibit differential sensitivity to both agonists and antagonists [37], suggesting that the divergence of AR paralogues, like in the case of ERβ, is sufficient to confer variability in ligand selectivity. Receptor sequences in other lineages have diversified further due to additional rounds of genome duplication–rainbow trout (*Oncorhynchus mykiss*) expresses two paralogous forms of ARβ [38] as a result of a postulated genome duplication specific to salmonids [39]. Differences in the number of paralogues of genes encoding steroid receptors, as well as their sequence diversity, contributes to inter-species differences in

sensitivity. Indeed, in the case of ER-mediated biomarkers such as vitellogenin, sensitivity in response to exposure to 17α-ethinylestradiol has been shown to vary significantly in different species [25]. This concept can likely be extended to other nuclear receptors (NRs) and indeed to any receptor capable interacting with xenobiotics. For non-model organisms, the lack of readily available sequence data for such receptors hinders development of quantitative biomarker assays and functional studies of relevant proteins such as steroid hormone receptors. Due to the continuing cost reductions associated with next-generation sequencing, the most efficient means of generating such sequence data for non-model organisms is high-throughput mRNA sequencing (RNA-Seq), which provides data necessary for the rapid design and implementation of focused gene expression assays (qPCR) and custom microarrays [40,41] and enables protein functional studies to proceed quickly by providing coding sequence data required for cloning and in vitro expression [42].

In the present study, we sequenced messenger RNA (mRNA) from brain, liver and gonads of sexually mature male and female *M. fluviatilis* and prepared an annotated database of candidate protein-coding transcripts using a *de novo* assembly approach. The resulting transcriptome was found to be highly comprehensive, with full-length homologues present for the majority of genes predicted from the genome sequence of a taxonomically related teleost. Since one of our group's major research interests is environmental endocrine disruption, we analysed the rainbowfish transcriptome to identify NRs and steroid biosynthetic enzymes that could potentially be involved in responses to EDCs. The results are presented here alongside an overview of the transcriptome as a whole.

Results and Discussion

Sequencing, transcript assembly and annotation

Poly-A$^+$ messenger RNA from brain, liver and gonad tissue from one adult male and one adult female Murray-Darling rainbowfish was sequenced using the Illumina HiSeq 2000 platform. The use of a single individual of each sex was intended to reduce the likelihood of assembly errors due to intra-specific polymorphism. A total of 258,906,260 read pairs from all libraries (S1 Table) were subjected to quality control procedures and assembled using Trinity [43,44], yielding a total of 262,050 putative transcripts in 149,742 clusters (Table 1).

148,394 likely open reading frames (ORFs) were identified using a Markov model-based approach to predict candidate protein-coding regions with a minimum length of 300 nucleotides. Redundancy in the complete set of ORFs was reduced by clustering sequences based on similarity of the deduced protein sequences, resulting in 43,841 non-redundant ORFs corresponding to a deduced amino acid sequence identity of $\geq$ 95%. Of these, 25,339 were deemed complete in terms of the coding sequence (i.e. beginning with a start codon and ending with a

Table 1. Statistics of the transcriptome assembly.

Trinity transcripts	262,050
Clusters	149,742
GC content (%)	44.51
Contig N50	3,367
Median contig length	655
Mean contig length	1539
Mean read length after QC	85
Mean fragment size	137
Total assembled bases	403,186,007

doi:10.1371/journal.pone.0142636.t001

stop codon), with the remaining sequences partial at either the 5' or 3' ends. Deduced proteins were annotated using functional vocabularies associated with homologous sequences (HHblits e-value ≤ 1e-10) in the Swiss-Prot database [45], with assignment of putative functions derived from numerous sources including Gene Ontology [46], KEGG [47], and EggNOG [48]. Expression levels for each non-redundant ORF were estimated by estimating abundance according to read coverage as described in the methods section. The complete set of assembled transcripts along with JAMp-derived functional annotations and normalised expression data for the non-redundant protein-coding transcripts are available through a web interface (http://annotation.insectacentral.org/waite/). Assembled transcripts and raw reads are available from NCBI Bio-Project (http://www.ncbi.nlm.nih.gov/bioproject/), accession PRJNA282059. Lists of assembled transcripts, predicted coding regions and non-redundant deduced protein sequences are also available from the CSIRO Data Access Portal (http://dx.doi.org/10.4225/08/563BDB4789194).

Assessment of the completeness of the transcriptome assembly

For the present study, we did not expect to identify a complete set transcripts spanning the entire genome due to the limited number of tissues sampled and the fact that we only sampled adult specimens maintained under laboratory conditions. However, we were interested in estimating the completeness of the assembly in terms of both the number of non-redundant protein-coding sequences present in the rainbowfish transcriptome that could be considered as homologous to sequences from known complete eukaryotic genomes, and the degree of coverage present for each homologue. We approached these questions using two methods, the first utilising a limited subset of highly conserved eukaryotic proteins [49] and the second involving the complete predicted proteome of a close taxonomic relative.

Firstly, we identified rainbowfish sequences corresponding to a subset of 'core eukaryotic genes' (CEGs) using the CEGMA software package [49], which is commonly used to assess the completeness of draft genome sequence assemblies but has also been applied to transcriptome assemblies [50,51]. According to CEGMA analysis, 232 of 248 CEGs (93.55%) were present as complete sequences in the rainbowfish transcriptome, with at least partial matches present for 236 CEGs (95.16%). Because Japanese medaka (*Oryzias latipes*) is the closest relative of rainbowfish for whole-genome sequence data is available, we also used the medaka subset of CEG mappings (available from http://korflab.ucdavis.edu/Datasets/genome_completeness/) as queries to search for homologous sequences in the non-redundant set of predicted rainbowfish proteins using BLASTP [52]. Of the 421 medaka CEGs, 405 unique rainbowfish sequences exhibiting high similarity (expect value ≤ 1e-10) and minimum query coverage of 70% (referring to the length of the alignment as a proportion of the length of the query sequence) were obtained (96.2%; S2 Table). The mean query coverage of these matches was 95.83% and the mean sequence identity of the alignments was 90.4%. The query coverage distribution shown in Fig 1 (inset) suggested that the number of sequences exhibiting low query coverage, indicative of either partial ORFs or non-orthologous hits to highly conserved domains, was low. These data indicate that full-length homologues were present in the rainbowfish transcriptome assembly for the majority of CEGs mapped in the medaka genome.

Secondly, we identified rainbowfish protein sequences homologous (BLASTP e-value ≤ 1e-10) to the entire set of medaka proteins predicted from the *O. latipes* genome sequence (ftp://ftp.ensembl.org/pub/release-76/fasta/oryzias_latipes/pep/; Ensembl release 76). This revealed 23,601 rainbowfish sequences homologous to 24,674 medaka proteins (95.7%; S3 Table), with a mean query coverage 92.1%. Query coverage distributions are shown in Fig 1. These values indicate that despite the limited number of tissues sampled in the present study, the assembled
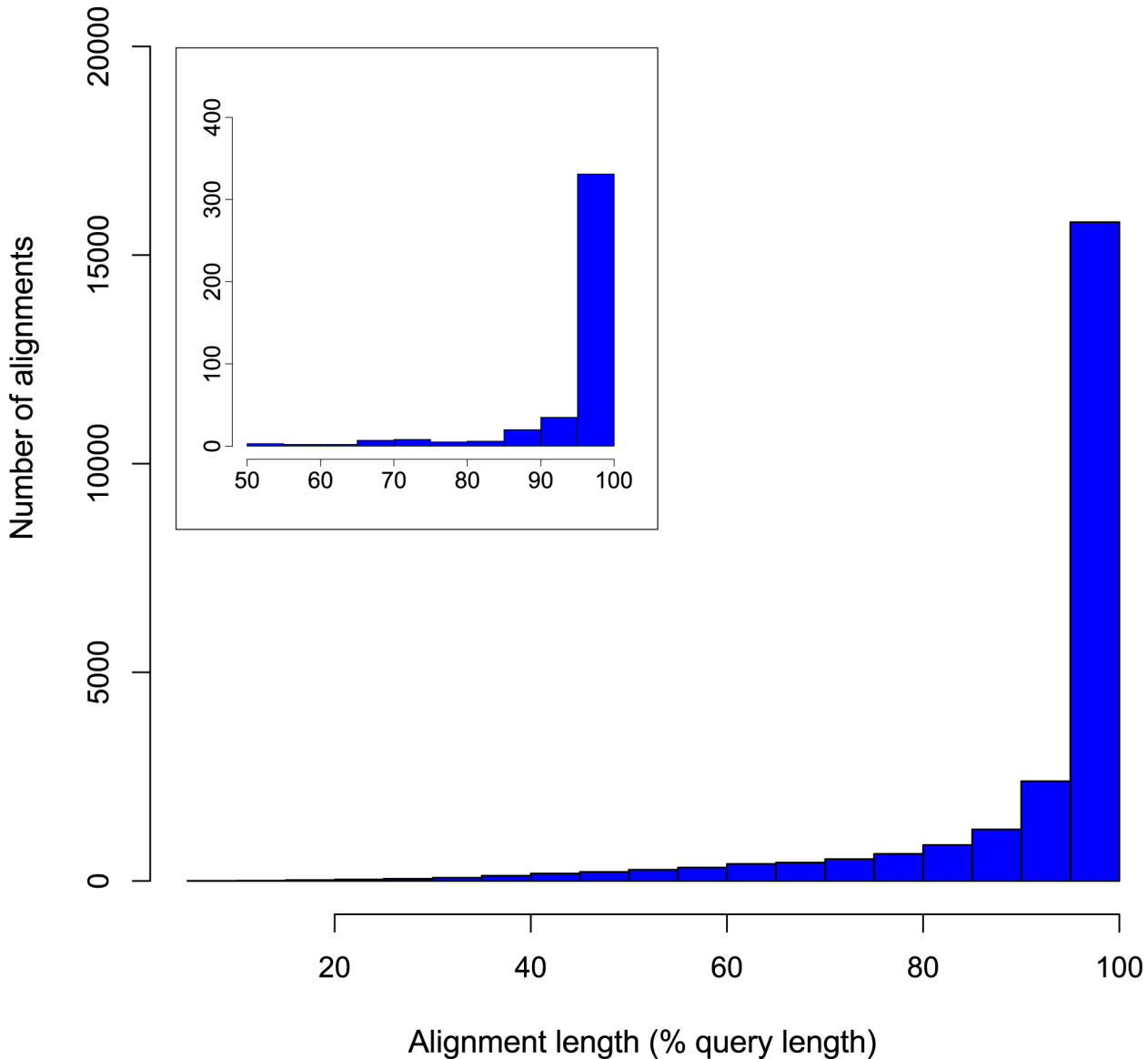
Fig 1. Analysis of the completeness of the rainbowfish transcriptome. Sequence similarity searches were conducted using the full set of predicted proteins derived from the Japanese medaka genome (*Oryzias latipes*; Ensembl release 76) as BLASTP queries against translated open reading frames from the Murray-Darling rainbowfish transcriptome. From 24,674 medaka query sequences, 23,601 high-quality (BLASTP e-value $\leq$ 1e-10) matches were obtained (95.7%) with a mean query coverage of 92.1%. Alignment length distributions are shown (bin size 5%), expressed as a proportion of medaka protein sequence lengths (% query coverage). **INSET:** A similar analysis conducted using the highly conserved subset of core eukaryotic genes (CEGs) from medaka.

doi:10.1371/journal.pone.0142636.g001

transcriptome contains full-length orthologues for a large proportion of the proteins predicted from medaka genome.

Identification of putative nuclear receptors in the rainbowfish transcriptome

The focus of the present study was to generate sequence data for key proteins of the reproductive endocrine system—primarily steroidogenic enzymes, nuclear receptors and the target genes thereof—to aid in the development of molecular methods for predicting of the potential

impacts on fish health of various endocrine-active contaminants. Since the best-studied EDCs are modulators of estrogen and androgen receptors, obtaining full-length sequence data for all subtypes of these receptors in rainbowfish was a high priority, but we were also interested in NRs that could potentially be activated or inhibited by a diverse range of environmental contaminants of emerging concern, including active pharmaceutical ingredients.

The major source of pharmaceuticals in surface waters is sewage, which contains a wide range of pharmaceutical drugs and metabolites that are incompletely removed during treatment [53,54,55,56,57,58,59,60,61]. Many drugs designed to target NRs in humans have the potential to interact with fish NRs due to the relatively high degree of sequence conservation amongst vertebrates. Indeed, some pharmaceuticals have been shown to affect developmental and reproductive endpoints in fish. Examples include dexamethasone, a steroidal glucocorticoid receptor (GR) agonist that inhibits growth and development in fathead minnow [62] and 19-nortestosterone-derived synthetic progestogens, which exhibit androgenic activity *in vivo* [63,64,65] and are potent agonists of ARs from fathead minnow [66] and Murray-Darling rainbowfish [67]. Peroxisome proliferator-activated receptors (PPARs) are also drug targets, with PPARα and β ligands including lipid-lowering fibrates, and anti-diabetes drugs that target PPARγ. While little data is currently available regarding the effects of pharmaceuticals on fish PPARs, fibrates have been shown to activate PPARα and β from rainbow trout [68], and the anti-diabetic drug rosiglitazone is a weak agonist of PPARγ from sea bream [69]. Because NRs control numerous fundamental developmental and physiological processes in vertebrates and are amenable to modulation by xenobiotics in the aquatic environment, the functional characterisation of NRs in fish species used for toxicological research will significantly advance our understanding of the potential risks to fish health posed by various endocrine-active environmental contaminants.

Canonical NRs are comprised of a number of functional domains necessary for their activity as ligand-activated transcription factors, including an amino-terminal activation domain, a DNA-binding domain (DBD), and a ligand-binding domain (LBD) (reviewed by Robinson-Rechavi et al., [70]). The most highly conserved regions are the DBD and the LBD, which hence can be used to annotate candidate NRs according to sequence similarity with known receptors. To identify putative NRs in the rainbowfish transcriptome, similarity searches were undertaken using the NCBI Batch CD-Search tool (http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi), which searches the NCBI Conserved Domain database to identify putative functional domains in query sequences. Rainbowfish sequences exhibiting high-scoring matches to the both the DBD superfamily (Pfam cl02596; 'NR_DBD_like') and the LBD superfamily (Pfam cl11397; 'NR_LBD') were identified, resulting in a list of 66 candidate nuclear receptors (S4 Table). This list was manually curated and annotated using NRs predicted from the Japanese medaka genome (listed in KEGG BRITE ola03310) resulting in a set of 61 candidate rainbowfish NRs (S5 Table). To categorise the NRs into subfamilies [71], a phylogenetic analysis was undertaken based on a multiple alignment of DBD and LBD sequences from the putative rainbowfish and medaka NRs (Fig 2). This revealed that representatives for all seven NR groups were present in the rainbowfish transcriptome, with the NR1, NR2 and NR3 subfamilies comprising the majority of sequences. The mean query coverage of rainbowfish candidate NRs with respect to medaka NRs was 87.3% and the mean amino acid identity was 85.9%.

The availability of sequence data for rainbowfish nuclear hormone receptors will allow us and other researchers to establish *in vitro* concentration-response relationships for various endocrine-active compounds using reporter gene transactivation assays. *In vitro* reporter gene assays have been demonstrated to correlate well with *in vivo* sensitivity to endocrine disruptors such as xenoestrogens [25,35]. Based on data generated as part of the present study, we have already successfully amplified full-length coding sequences for rainbowfish ARα and ARβ and
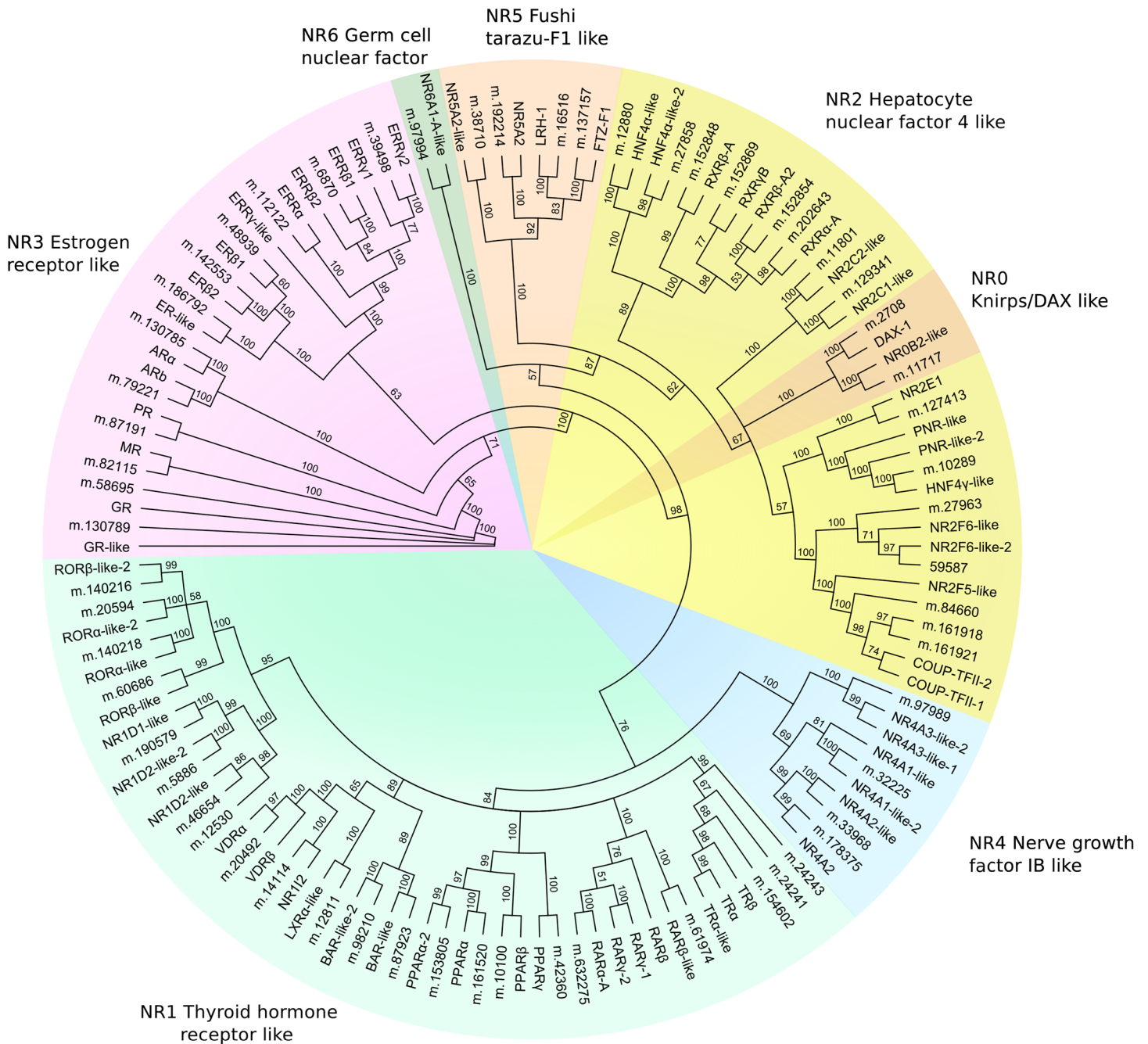
Fig 2. Phylogenetic analysis of candidate nuclear receptors from rainbowfish and nuclear receptors sequences from the Japanese medaka (*Oryzias latipes*) genome. Node labels indicate percent bootstrap support from 100 replications. Branch lengths are not to scale. Candidate nuclear receptors from the rainbowfish transcriptome are indicated by their coding sequence identifiers (m.*). Abbreviations: AR, androgen receptor; BAR, bile acid receptor; COUP-TF, chicken ovalbumin upstream promoter transcription factor; DAX-1, dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1; PNR, photoreceptor cell-specific nuclear receptor; ER, estrogen receptor; ERR, estrogen receptor-related; FTZ-F1, fushi tarazu factor 1; GR, glucocorticoid receptor; HNF, hepatocyte nuclear factor; LRH, liver receptor homologue; LXR, liver X receptor; MR, mineralocorticoid receptor; NR, nuclear receptor; PPAR, peroxisome proliferator-activated receptor; PR, progestin receptor; RAR, retinoic acid receptor; ROR, RAR-related orphan receptor; RXR, retinoid X receptor; TR, thyroid hormone receptor; VDR, vitamin D receptor. NRs not assigned a common name are denoted by their subfamily, group and member number, i.e. 'NR5A2' represents nuclear receptor subfamily 5, group A, member 2.

doi:10.1371/journal.pone.0142636.g002

confirmed their function as androgen-dependent transcription factors capable of transactivation through canonical androgen response elements [37]. Our *in vitro* study showed that rainbowfish ARα and ARβ exhibit markedly different sensitivities to both synthetic androgens and anti-androgens, indicating that the divergence of paralogous duplicates arising from the TSGD is sufficient to confer differential responses to selected ligands. We have also found that synthetic progestins derived from 19-nortestosterone are potent agonists of rainbowfish ARα and ARα but do not activate the rainbowfish progestin receptor [67], adding to recent data indicating that fathead minnow AR is activated strongly by progestins [66] and providing further mechanistic evidence that androgen receptor activation underlies the effects observed in various species exposed to progestins *in vivo* [63,64,65,72,73].

## Identification of candidate steroidogenic enzymes

Normal functioning of endocrine systems in fish can be altered by EDCs not only by direct interaction with hormone receptors, but also via the binding and inhibition of steroidogenic enzymes [74], most of which belong to the cytochrome P (CYP)450 superfamily of haem-containing mono-oxygenases. One of the best studied examples in the context of EDCs is the inhibition of the enzyme responsible for conversion of androgens to estrogens, CYP19 aromatase, by azole compounds including fungicides [75]. Mechanistic studies using the specific aromatase inhibitor, fadrazole, have shed light on the dramatic effects aromatase inhibitors can elicit in fish [76]. Exposure to fadrazole during a developmental window critical for sexual differentiation can result in the complete masculinisation of genetically female chinook salmon [77]. Teleosts may be particularly susceptible to aromatase inhibitors due to presence of an additional genetic locus encoding a brain-specific subtype denoted CYP19A1b, expressed at high levels during early development [78].

Changes in the expression of steroid biosynthetic genes can also occur in response to EDCs and is considered one of the causative mechanisms underlying altered plasma steroid levels in exposed organisms [74,79,80,81,82]. Specific examples include the induction of aromatase gene expression in response to estrogens [79,83] and increased expression of 20β-hydroxysteroid dehydrogenase (20β-HSD), a key enzyme in the synthesis of the progestins in fish, in response to anti-androgens [84].

To identify candidate steroid biosynthetic enzymes in the rainbowfish transcriptome, *O. latipes* proteins annotated as members of steroidogenic pathways (KEGG pathway ola00140 'steroid hormone biosynthesis') were used as query sequences for similarity searches of the complete set of translated rainbowfish ORFs. This resulted in a list of 41 unique high-quality matches (maximum e-value of 2.8e-32) with a mean query coverage of 92.46% (S6 Table). At least one full-length match was present for each of the major enzymes involved in steroidogenesis in teleosts. To demonstrate that the majority of key enzymes involved in steroid hormone biosynthesis were identified in the present study, rainbowfish ORF identifiers corresponding to the best match for each enzyme are listed alongside enzyme descriptions in a schematic diagram of steroid biosynthesis in teleosts (Fig 3; adapted from Häggström and Richfield [85] with reference to pathways described in [86,87,88]).

The presence of a 5α reductase in Murray-Darling rainbowfish transcriptome raises the question of whether the androgen 5α-dihydrotestosterone (DHT) may be present at physiologically relevant levels in this species, as has been reported for fathead minnow (*Pimephales promelas*) [87,89]. We have observed using in vitro reporter assays that DHT is comparable in potency with 11-KT as an agonist of both ARα and ARβ [37], suggesting that endogenous DHT, if present, may be an important androgen in *M. fluviatilis*.
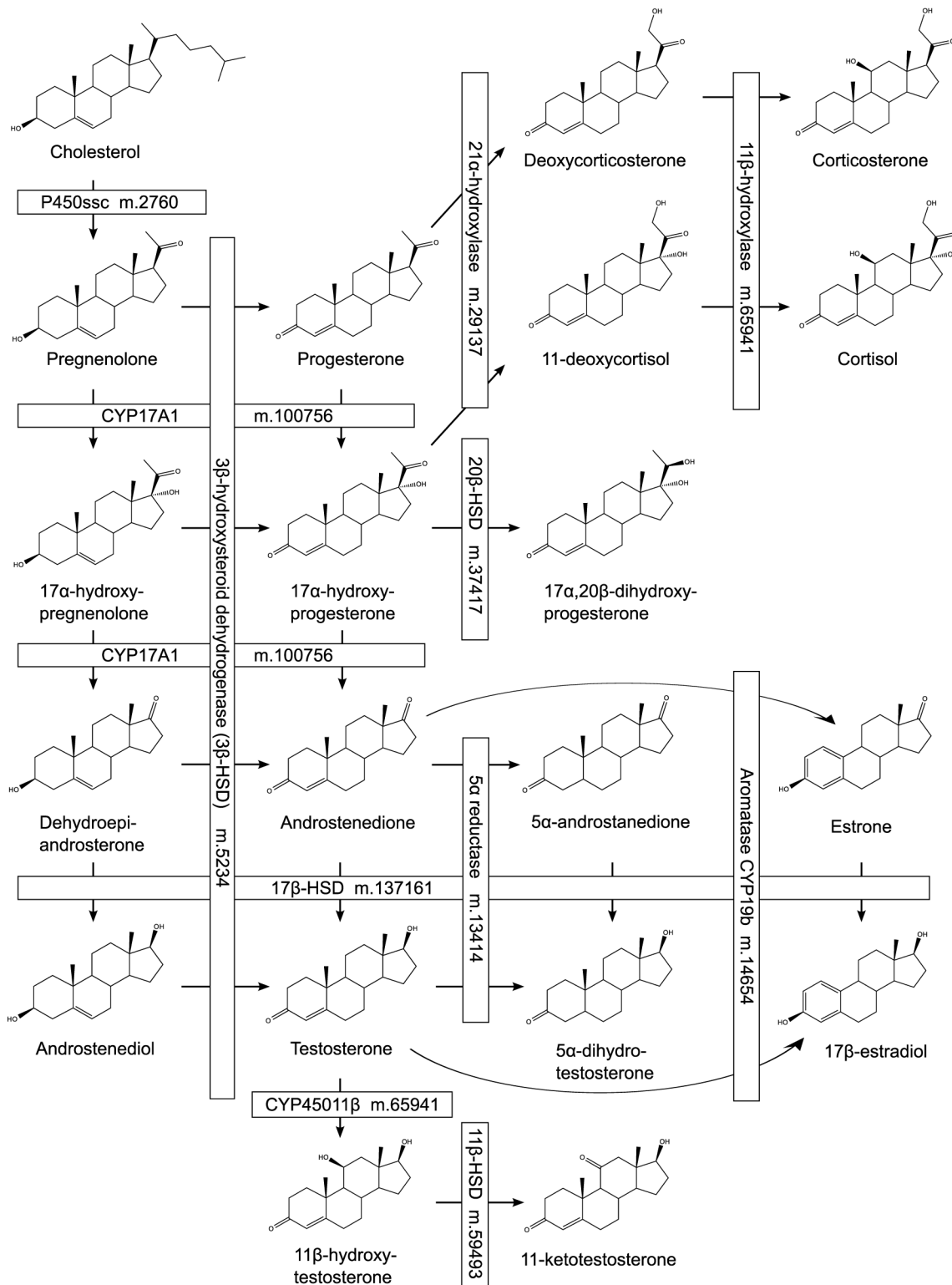
Fig 3. Schematic diagram of likely steroidogenic pathways in Murray-Darling rainbowfish. ORF identifiers corresponding to deduced proteins displaying high similarity to biosynthetic enzymes are shown alongside enzyme descriptions at each step of the biosynthetic pathway. Adapted from Häggström and Richfield [85] with reference to known teleost steroidogenic pathways [86,87,88].

doi:10.1371/journal.pone.0142636.g003

Identification of transcripts for estrogen-responsive genes

A number of genes known to be positively regulated at least in part by ERs have been validated as biomarkers of (xeno)estrogen exposure in male or juvenile fish [90,91]. Vitellogenin (VTG) and choriogenins (ChgH and ChgL), precursors to yolk and egg envelope proteins, respectively, are the most widely employed biomarkers of ER activation [91,92]. Plasma VTG levels are often determined using specific antibodies, but since the site of VTG synthesis is the liver, hepatic mRNA levels for VTG are closely correlated with plasma protein and can be utilised as a reliable surrogate if antibodies for a particular species are not available. Expression of the *esr1* gene in fish (encoding ERα) is also induced in response to estrogens [93,94,95], indicating an auto-regulatory mechanism. ERα mRNA is also used to infer exposure to (xeno)estrogens in male fish, since increased levels have been positively correlated with histological changes consistent with intersex condition [24]. We have previously developed quantitative reverse-transcription PCR (qRT-PCR) assays for selected ER-regulated transcripts in rainbowfish, including VTG [16], choriogenin L (ChgL [18]), and ERα [16], but to date only ERα has been fully sequenced for this species.

To identify putative estrogen-responsive genes in the rainbowfish transcriptome, we used a subset of estrogen-responsive genes from medaka as BLAST queries. Full-length sequences were found for three VTG subtypes (denoted VTG A, B and C), ERα, ERβ1, ERβ2, choriogenin H (ChgH) and ChgL (Table 2). Multiple putative splicing variants of similar length were assembled for VTG A and B. ERβ1 and ERβ2 sequences required manual curation in order to obtain full-length putative coding sequences. In the case of ERβ1, an apparent frameshift was present in the predicted coding sequence. In an attempt to resolve a complete coding sequence, we generated an alternative assembly using only ovary-derived reads, which was found to contain an ERβ1-like ORF of similar length to published sequences from other teleosts. For ERβ2, a suspected assembly error was apparent in a T-rich region, resulting in a frameshift in the predicted coding sequence. This error was confirmed by sequencing a PCR product generated

Table 2. Rainbowfish transcripts with high similarity to estrogen-responsive genes from related teleosts.

Description	Rainbowfish transcript|ORF	Deduced protein length (aa)	Closest BLAST match (GenBank non-redundant protein sequences, *M. fluviatilis* omitted)	Length of closest match (aa)	GenBank Accession
Vitellogenin A	comp98304_c0_seq1|m.58126, comp98304_c0_seq2|m.58130, comp98304_c0_seq3|m.58134	1,717, 1,699, 1,690	vitellogenin [*Gambusia affinis*]	1,695	BAD93697
Vitellogenin B	comp102542_c3_seq1|m.120426, comp102542_c3_seq2|m.120431, comp102542_c3_seq3|m.120436, comp102542_c3_seq4|m.120441	1,733, 1,717, 1,713, 1,709	vitellogenin Ab [*Dicentrarchus labrax*]	1,692	AFA26670
Vitellogenin C	comp42500_c0_seq1|m.2380,	1,244	vitellogenin C [*Dicentrarchus labrax*]	1,275	AFA26671
Estrogen receptor α	comp105150_c0_seq9|m.186792	611	estrogen receptor alpha [*Odontesthes bonariensis*]	602	ABY19510
Estrogen receptor β1	transcript from alternative assembly (ovary)	559	estrogen receptor beta 1 [*Odontesthes bonariensis*]	558	ABY19511
Estrogen receptor β2	comp103476_c2_seq6|m.142553 (manually edited to remove likely assembly error in poly-U tract and verified by sequencing)	665	estrogen receptor beta 2 [*Oreochromis niloticus*]	666	NP_001266406
Choriogenin H	comp78177_c1_seq1|m.9381	466	choriogenin H minor [*Fundulus heteroclitus*]	452	BAJ07537
Choriogenin L	comp78177_c1_seq1|m.9382	440	L-SF precursor [*Oryzias latipes*]	420	NP_001098273

doi:10.1371/journal.pone.0142636.t002

using primers spanning the CDS, and a complete coding sequence was thus btained. Full-length coding sequences for estrogen-responsive transcripts generated in the present study will facilitate further gene expression and functional studies.

Identification of ERα transcript variants

Alternative splicing of mRNA resulting in exon skipping or variable first or last exons generates protein isoforms with altered function compared to the canonical splice form (reviewed by Black et al. [96]). It has been estimated from high-throughput sequencing that approximately 95% of multi-exon genes in the human genome undergo alternative splicing [97], suggesting that the spectrum of protein products encoded by higher eukaryotic genomes greatly outnumbers the gene count. The occurrence of multiple ERα transcript variants derived from alternative splicing of pre-mRNA has been reported in a number of teleosts [98,99,100,101,102,103,104,105,106] and can contribute to environmental adaptations. For example, in killifish adapted to high levels of xenoestrogens in a contaminated waterway, alternative mRNA splicing has been proposed as one mechanism by which estrogenic responses are dampened [98]. RNA-Seq is a powerful tool for discovery of alternatively spliced transcripts [97].

According to comparative studies, Trinity is the leading *de novo* transcript assembly software for the reconstruction of mRNA splicing variants [107,108,109], prompting us to investigate whether rainbowfish ERα mRNA splicing variants were accurately predicted by Trinity in the present study. Sequences in the Trinity transcript cluster *comp105150_c0* exhibited high similarity to teleost ERα coding sequences and contained 12 transcripts with ORFs of varying lengths due to disruption frameshifts arising from apparent alternative splicing or the presence of unspliced introns. Possible donor and acceptor dinucleotides [110] could be identified in those variants with large apparent inserts (data not shown) suggesting the presence of introns, possibly attributable to the co-purification of immature RNA in the samples used for sequencing.

Three of the transcripts, *comp105150_c0_seq9*, *comp105150_c0_seq11* and *comp105150_c0_seq12*, encoded putative full-length ORFs typical of teleost ERα and were therefore chosen for further study. Interestingly, all putative transcripts in the cluster contained identical 5' ends, suggesting that exon skipping may underlie the apparent transcript diversity in the 5' region–*seq9* and *seq11* differed from *seq12* by the inclusion of additional sequence following the first exon, with a further insertion predicted for *seq9* relative to that predicted for *seq11*. The protein encoded by *seq9* and *seq11* differs from the *seq12*-encoded protein by only two amino acids at the amino terminus, ('M**LL**RQS. . .' in Seq9 and 11 and 'M**FK**RQS. . .' in Seq12). Interestingly, these sequences are consistent with two conflicting medaka *esr1* gene models, one of which (GenBank AB033491.1) predicts a structure similar to *comp105150_c0_seq12*, while the other (ENSORLG00000014514, Ensembl medaka release 78) predicts an exon structure and CDS similar to *comp105150_c0_seq9* and *seq11* (with the exception of putative first exon, which is omitted in the Ensemble gene model, with exon 1 annotated from the start of the CDS). Like the predicted rainbowfish proteins, translation of two medaka coding sequences yields two proteins differing only in first few amino acids—'M**LL**RQS. . .' in the longer form (UniProt H2MHS4, identical to the translated CDS from ENSORLG00000014514) and 'M**SK**RQS. . .' in the shorter form (UniProt P50241, identical to the translated CDS from GenBank AB033491.1). Interestingly, two similar ERα variants have been proposed in guppy (*Poecilia reticulata*), with deduced amino acid sequences differing by only two residues ('M**LL**RQN. . .', GenBank AB621910; and 'M**YK**RQN. . .', GenBank XP_008395912).

ERα splicing variants sharing a common first exon have also been described in killifish (*Fundulus heterclitus*) and have been postulated to result from the splicing of exon 1 to two alternative second exons denoted 2a or 2b, which differ in length by approximately 200 bases

[98]. The longer killifish ERα transcript is similar to the *comp105150_c0_seq9-seq11* transcript described in the present study. The shorter killifish ERα transcript, first described by Urushitani et al. [111] and later confirmed by Greytak and Callard [112], contains an ORF of identical length to that in the longer transcript but encodes a protein that, similar to the putative rainbowfish isoforms and the medaka gene models, differs by only a few amino acids immediately following the initiator methionine ('M**LLR**Q...' and 'M**YKG**Q...' from the long and short transcripts, respectively). Both the long and the short killifish transcripts have two possible translational start codons–the second located 141 bases downstream from the first, which would result in the omission of 47 amino acids from the amino terminus. Translation from the second ATG of either transcript isoform would result in a receptor similar to the short form of rainbow trout ERα (ERα$_S$), which displays higher ligand-independent transactivation activity than the long form (ERα$_L$) [100], due to the lack of a possible 'suppression' domain.

To confirm expression of the predicted rainbowfish ERα transcripts, we performed a series of end-point (non-quantitative) PCR reactions using primers targeting specific transcript variants (Fig 4A and Table 3). PCR products for the coding regions of both *seq9-seq11* and *seq12* were successfully amplified (Fig 4B) confirming that these coding variants were expressed to some degree. Both transcripts were expressed almost exclusively in female liver, with the *seq12* variant expressed at higher levels than *seq9-seq11* (although it should be noted that no biological replication was used in the present study and these results are from one male and one
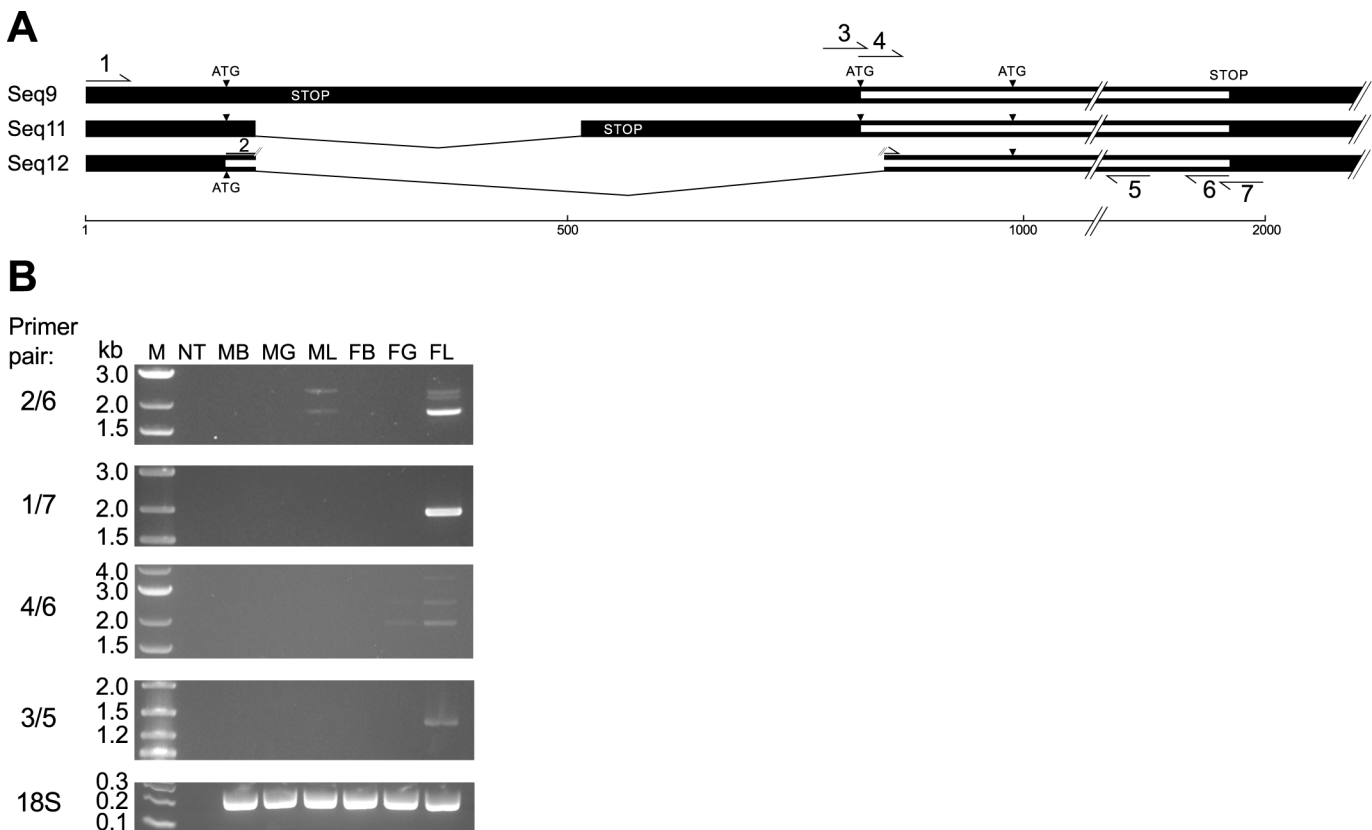
Fig 4. Confirmation of selected transcript variants of a putative rainbowfish ERα cluster by endpoint reverse-transcriptase PCR (RT-PCR). (A) Schematic diagram of three transcripts in the cluster–*comp105150_c0_seq9*, *comp105150_c0_seq11* and *comp105150_c0_seq12*, showing binding sites for the PCR primers listed in Table 4. (B) Agarose gel electrophoresis of RT-PCR products generated using the indicated primer pairs. Primers targeting 18S rRNA were used to confirm the quality of cDNA in all samples. Abbreviations: kb, kilobases; M, marker; NT, no-template control; MB, male brain; MG, male gonad; ML, male liver; FB, female brain; FG, female gonad; FL, female liver; 18S, 18S rRNA.

Table 3. Primers used for RT-PCR analysis of rainbowfish ERα 5'-end variants.

Primer number	Target transcript	Sequence (5'-3')
1	Seq9/11/12	CACCATGATTATTGATTCGGC
2	Seq12	ATGTTTAAGAGGCAGAGCCTG
3	Seq9/11	GCTGTGATGTTGCTCAGGCAG
4	Seq9/11	ATGTTGCTCAGGCAGAGCCT
5	Seq9/11/12	GCAGAAAGAAAAGGCACCAG
6	Seq9/11/12	TCATAGTGCGTGGGCGCA
7	Seq9/11/12	CCATCTCATAGTGCGTGGGC

doi:10.1371/journal.pone.0142636.t003

female individual fish). Additional bands appear at higher molecular weight may correspond to immature mRNA containing unspliced introns apparent in other sequences in the cluster (*seq1-8* and *seq10*). The solitary band resulting from primer pair 1/7 indicates that the *seq12* transcript is the predominant isoform. In fact, alignment of the rainbowfish transcript variants with the medaka *esr1* gene sequence (S1 Fig) shows that rainbowfish *seq9* may represent immature RNA, since the entire 5' UTR aligns well with the medaka genome sequence, while the predicted *seq11* may constitute an shorter transcript derived from exon skipping as predicted for *esr1* in the draft guppy genome (Entrez Gene ID: 103457510). No short variants similar to the killifish ERα variants derived from exon skipping [98] were detected using RNA-Seq or RT-PCR in the present study.

Transcripts encoding CYP450-family enzymes

In addition to their role in steroidogenesis, CYP450-family enzymes participate in the oxidative metabolism of a wide variety of endogenous and exogenous chemicals. Vertebrate CYP450s are classified into 18 families with varying substrate specificity [113]. Members of families CYP1, CYP2 and CYP3 are notable for their ability to metabolise xenobiotics such as polycyclic aromatic hydrocarbons (PAH)s, dioxins and polychlorinated biphenyls (PCBs). Elevated expression of CYP1A subfamily isoforms in fish liver tissue is a widely utilised biomarker for this class of environmental contaminants (reviewed by Bucheli and Fent [114]). The zebrafish (*Danio rerio*) genome contains 94 genes spanning the 18 vertebrate CYP450 families, including an apparent expansion of the CYP2 family to 47 members compared to 16 members in humans, with many of the zebrafish genes present on duplicated genomic regions arranged in tandem [115]. The genome of the pufferfish (*Takifugu rubripes*) encompasses 54 CYP450 genes from 17 families, including at least 24 putative CYP2-family sequences [116].

To identify full-length rainbowfish transcripts encoding putative CYP450-family enzymes, we retrieved zebrafish CYP450 protein sequences from the Ensembl zebrafish genome assembly Zv9 associated with PFAM family PF00067 (cytochrome P450) and performed a BLASTP similarity search against the complete list of predicted rainbowfish proteins. This resulted in 52 unique hits with at least one high-scoring match (maximum e-value for all hits was 3.06e-25) present for each of the 18 vertebrate CYP450 families (S7 Table). The mean query coverage was 90.76%. Like other teleosts, the CYP2 family had the largest number of representatives in the rainbowfish transcriptome, with 15 putative coding sequences. Given the high number of CYP2-family genes present in fugu [116] and zebrafish [115], it is possible that additional rainbowfish CYP2 transcripts were not detected in the current study due to low-level expression under normal laboratory conditions or in the tissues sampled. A candidate rainbowfish ORF encoding a protein with 70.6% amino acid identity to zebrafish CYP1A and query coverage of

98.85% was identified (*comp94347_c0_seq1|m.33721*), which may be suitable for use as a biomarker of compounds that induce the adaptive stress response.

Transcripts involved in the xenobiotic response pathway

PCBs, PAHs and dioxins can induce potent xenobiotic defence responses in fish by binding to and activating the aryl hydrocarbon receptor (AhR). The functionality of AhR is unique in that it is the only ligand-activated member of the Per-ARNT-Sim family of basic-helix-loop-helix (bHLH) transcription factors [117,118]. Ligand-bound AhR is translocated to the nucleus, interacting with the Ah receptor nuclear translocator (ARNT) to induce the expression of target genes such as CYP1A by binding to dioxin response elements located in their proximal promoter regions [119]. The number and diversity of AhR genes in teleosts is considerable, with numerous subtypes of two distinct AhR genes denoted AhR1 and AhR2 expressed by some species [120,121,122]. For example, Fugu (*Takifugu rubripes*) expresses two AhR1 subtypes (*ahr1a* and *1b*), and three AhR2 subtypes (*ahr2a*, *2b*, and *2c*) [121,123]. In zebrafish, AhR1A is not capable of activating transcription through known response elements and thus may not be functional, while AhR1B and AhR2 are functional [123]. Zebrafish AhR2 binds TCDD and transactivates genes via AhR-responsive elements, while AhR1B is capable of transactivation but may be constitutively active since it does not appear to be TCDD-responsive [123].

To identify candidate members of the AhR pathway in the rainbowfish transcriptome, deduced protein sequences associated with the GO term 'aryl hydrocarbon receptor activity' (GO: 4874) were extracted from the JAMp database using the keyword search function in the web interface. Putative descriptions were inferred by performing individual BLASTP searches against the GenBank non-redundant protein database (S8 Table). The list of rainbowfish proteins included a partial ORF similar to teleost AhR1, a complete ORF encoding a protein with high similarity to teleost AhR2 sequences, and a number of full-length ORFs with high similarity to various isoforms of teleost ARNT (Table 4). These sequences could be useful for designing gene expression assays for biomarkers related to xenobiotic metabolism in Murray-Darling rainbowfish exposed to environmental contaminants such as dioxins, PCBs and PAHs, while the full-length coding sequence for an AhR2A homologue could be used to develop xenobiotic-responsive reporter gene assays capable of screening environmental samples for such contaminants.

Conclusions

Here we describe the sequencing, assembly, analysis and annotation of the Murray-Darling rainbowfish protein-coding transcriptome from adult male and female brain, liver and gonads. The rainbowfish transcriptome contains full-length orthologues for more than of 95% of the proteins predicted from the medaka genome, a taxonomically related species. The assembly therefore is likely to comprise not only a comprehensive repertoire of transcripts involved in steroid hormone biosynthesis and activity, but also the vast majority of transcripts expressed in adult liver, brain and gonads. This represents a valuable resource for researchers interested in protein structure and function, molecular toxicology, contaminant-response genomics and population genetics of the *Melanotaenia* genus, which is widely distributed throughout Australasia. It is becoming evident that genetic variability underlies the observed differences in inter-species responses to endocrine-active environmental contaminants. Thus, in aquatic environments impacted by such contaminants the responses observed in a single teleost species cannot be assumed to be representative of the potential impacts on fish in general and it may be more appropriate to characterise responses in numerous endemic species. Methods for rapidly obtaining sequence data for the nuclear receptors that largely define these interspecies differences will undoubtedly assist research efforts in this area. The methodology used in the

Table 4. Transcripts involved in aryl hydrocarbon receptor (AhR)-mediated adaptive stress responses. Transcripts were classified into two functional groups–the nuclear AhR receptor complex (GO: 0034753) or AhR target genes comprising selected transcripts strongly upregulated by 2,3,7,8-Tetrachlorodi-benzo-p-dioxin selected from Li et al. [124] and Watson et al. [125].

Putative description	Transcript/CDS identifier	Deduced protein length (aa)	Closest match (GenBank non-redundant protein sequences, *M. fluviatilis* omitted)	Length of closest match (aa)	Accession
AhR receptor complex					
Aryl hydrocarbon receptor 1A (AhR1A)	comp328301_c0_seq1|m.267373	268	aryl hydrocarbon receptor 1 [*Fundulus heteroclitus*]	944	AAR19369
Aryl hydrocarbon receptor 2A (AhR2A)	comp94961_c0_seq1|m.36116	914	aryl hydrocarbon receptor 2 [*Pagrus major*]	990	BAE02825
Aryl hydrocarbon receptor nuclear translocator (ARNT)	comp106403_c1_seq1|m.227926	638	PREDICTED: aryl hydrocarbon receptor nuclear translocator-like protein 1-like isoform X2 [*Oreochromis niloticus*]	638	XP_005449385
Aryl hydrocarbon receptor repressor (AhRR)	comp89767_c1_seq1|m.20660	639	PREDICTED: uncharacterized protein LOC100696058 [*Oreochromis niloticus*]	694	XP_003460145
AhR target genes					
Cytochrome P450 1A (CYP1A)	comp94347_c0_seq1|m.33721	421	Cytochrome P450 1A [*Poecilia vivipara*]	521	AFN02446
Cytochrome P450, family 1, subfamily B, polypeptide 1 (CYP1B1)	comp83107_c0_seq4|m.12292	583	PREDICTED: cytochrome P450 1B1-like [*Stegastes partitus*]	537	XP_008278388
Cytochrome P450, family 1, subfamily D, polypeptide 1	comp81594_c0_seq2|m.11394	537	PREDICTED: cytochrome P450 1A1-like [*Poecilia formosa*]	535	XP_007564479
Stromal cell-derived factor 2 (SDF2)	comp88906_c0_seq1|m.19082	245	PREDICTED: stromal cell-derived factor 2-like [*Neolamprologus brichardi*]	226	XP_006784901
Heat-shock protein A5 (HSPA5) / Glucose-regulated protein, 78 kDa (GRP78)	comp99062_c0_seq1|m.65001	564	Glucose-regulated protein 78 [*Paralichthys olivaceus*]	654	ABG56392
Aldehyde dehydrogenase 3 family, member A1 (ALDH3A1)	comp101664_c3_seq2|m.102761	532	PREDICTED: aldehyde dehydrogenase, dimeric NADP-preferring-like [*Stegastes partitus*]	509	XP_008297392
Nuclear factor erythroid 2-related factor 2 (Nrf2)	comp93427_c0_seq1|m.30054	617	PREDICTED: nuclear factor erythroid 2-related factor 2 [*Stegastes partitus*]	613	XP_008304424
NADPH:quinone oxidoreductase 1 (NQO1)	comp95672_c0_seq1|m.40009	272	NAD(P)H dehydrogenase quinone 1 [*Fundulus heteroclitus*]	277	ABD43173
Fibroblast growth factor 13 (FGF13)	comp105488_c2_seq1|m.196118	247	PREDICTED: fibroblast growth factor 13-like isoform X1 [*Oreochromis niloticus*]	246	XP_003445268

doi:10.1371/journal.pone.0142636.t004

present study is generally applicable and adds to the body of work demonstrating of the efficacy of RNA-Seq and *de novo* assembly in the rapid generation of protein-coding sequence data.

Materials and Methods

Animals

Sexually mature male and female *M. fluviatilis* were obtained from a commercial supplier (Aquarium Industries, NSW, Australia). Fish were maintained at 22°C under 16 h light / 8 h darkness photoperiod and fed dried fish flakes. The study was conducted with the approval of the Animal Ethics Committee of the Commonwealth Scientific and Industrial Research Organisation's Division of Food and Nutritional Sciences (reference #774-12/13). One male (2.91 g, fork length 65 mm) and one female (5.1 g, fork length 80 mm) were anaesthetised using MS-222 prior to euthanasia by decapitation. Brain, liver and gonads were excised immediately and snap-frozen in liquid nitrogen.

RNA isolation and sequencing

Total RNA was purified from approximately 15 mg liver or gonad tissue using the RNeasy® Mini Kit (QIAGEN Pty. Ltd., Australia). Total RNA from whole brain was purified using the RNeasy® Lipid Tissue Kit (QIAGEN Pty. Ltd., Australia). On-column DNase treatments were included during purification. Total RNA concentration and purity were estimated using a NanoDrop 1000 spectrophotometer (Thermo Scientific). RNA integrity was confirmed using the Bioanalzyer 2100 (Agilent).

Sequencing was undertaken by a service provider (Australian Genome Research Facility, Melbourne, Australia). Library preparation was performed using Illumina TruSeq reagents and sequencing was conducted using the HiSeq2000 platform. Briefly, RNA quality was determined using a Bioanalyzer 2100 (Agilent) before isolation of polyA$^+$ mRNA, fragmentation, and construction of cDNA libraries using the Illumina TruSeq kit according to the standard protocol. 100-cycle paired-end sequencing was undertaken using two lanes of an Illumina HiSeq2000 instrument, with three libraries multiplexed per lane.

Transcript assembly and annotation

For each library, average per-base quality and positional base composition were determined using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Bases with phred +33 quality scores of below 10 were trimmed from the 3' end of each read using Trimmomatic [126]. Bases displaying obvious compositional bias (this comprised the initial 14 bases of each read according to summary graphs) were removed from all reads. Despite reports that compositional bias at the beginning of Illumina reads can be attributed to unexpected weighting of base composition in priming hexamers [127], we opted for a conservative approach since the overall sequencing quality and yield were high.

Trimmed reads were assembled into putative transcripts using Trinity [44] (http://trinityrnaseq.github.io/), release r2012-06-08. Default values were used for the assembly parameters [43], with the exception that the '—min_kmer_cov' option defining the kmer coverage required for extension of contig ends was set to 2 in order to prevent contig extension by singleton kmers. 300 gigabytes of memory was allocated to the program, which was run using a high-memory server at the BRAEMBL high-performance computing facility, 'Barrine' (University of Queensland, St Lucia, Australia). To estimate fragment size of the paired-end libraries, reads were aligned against assembled transcripts using the default alignment options in RSEM (http://deweylab.biostat.wisc.edu/rsem/), with fragment size distributions visualised using the included R script.

Protein-coding regions were predicted using TransDecoder (http://transdecoder.github.io/) with the default parameters. TransDecoder utilises a Markov model trained on the longest open reading frames (ORFs) to predict a set of likely protein coding sequences. In cases where assembled transcripts contained more than one ORF, all ORFs were retained with the exception of those occurring completely within a larger ORF. To calculate read counts for predicted protein-coding regions, redundancy was reduced by clustering the full set of deduced amino acid sequences using CD-HIT [128,129], with a similarity cut-off of 95%.

Deduced protein sequences were annotated with the software Just_Annotate_My_proteins (JAMp; http://jamps.sourceforge.net/), a workflow that exploits ontologies associated with entries in the Swiss-Prot database (the reviewed subset of the UniProt Knowledgebase/UniProtKB) to annotate new protein sequences based on similarity to high-quality, manually curated proteins present in the Swiss-Prot database. The JAMp workflow has been described previously [42]. Briefly, JAMp uses HHblits [130] to search hidden markov models (HHMs) derived from a set of unknown proteins against HMMs derived from Swiss-Prot database, which provides greater

sensitivity than conventional similarity-based searches. Up to 10 HHblits alignments that met similarity criteria (homology probability $\geq$ 80%; HHblits score $\geq$ 70; homology e-value $\leq$ 1e-10; homology p-value 1e-12) were used for functional annotation of predicted rainbowfish proteins. To assign putative functions to predicted proteins, controlled vocabularies associated with matching UniProtKB entries were linked with a rainbowfish sequence only when experimental evidence for functional terms existed (annotations derived from sequence similarity alone were not considered). Controlled vocabularies used for functional annotation included Gene Ontology [46], EggNOG [48], and KEGG [47] and subsets of these databases. Functional annotations, ORFs and transcripts were stored in a PostgreSQL database for access via a web browser-based viewer built with ExtJS4 and canvasXpress.

Additional functional annotations were inferred by sequence similarity to other teleost sequences using BLAST searches [52,131], using controlled vocabularies obtained from the Gene Ontology database via BLAST2GO software [132] or extracted from matching sequences in the NCBI conserved domain database (CDD) using reverse position-specific BLAST [133].

Completeness of the transcriptome was assessed using CEGMA v2.4 [49] (which depends on blast+ v2.2.25, geneid v1.4.4, genewise v2.2.4 and hmmer v3.0) or by sequence similarity searches in amino acid space conducted using BLASTP. For the BLASTP searches, we used *O. latipes* CEGs from http://korflab.ucdavis.edu/Datasets/genome_completeness/ and the entire predicted proteome derived from the genome assembly from Ensembl release 76 (ftp://ftp.ensembl.org/pub/release-76/fasta/oryzias_latipes/pep/). Query coverage histograms were prepared with R using the built-in Stats package [134].

Geneious v7 was used to prepare alignments, for phylogenetic analyses, as a front-end for sequence similarity searches and for general curation and review of sequence data.

Confirmation of ERα transcript variants

Total RNA (1 μg) from brain, liver and gonads from male and female fish was reverse transcribed using the Quantitect® kit (Qiagen Pty. Ltd., Australia) according to the manufacturer's protocol, except that the incubation time for cDNA synthesis was extended to 1 h. PCR reactions (20 μL) contained 0.4 units Phusion® HotStart II polymerase (Thermo Fisher Scientific, Australia), 1 × HF buffer (Thermo Fisher Scientific, Australia), 0.5 μM dNTP mix (New England Biolabs, Inc.), 0.5 μM each primer and 1 μL of ten-fold diluted cDNA.

End-point PCR was conducted using Phusion HotStart II polymerase with 35–40 cycles. Optimal annealing temperatures for PCR reactions conducted using were calculated as recommended by the manufacturer (www.thermoscientificbio.com/webtools/tmc/). Extension times used were 30 s per kb of the largest expected product. Products were separated on 1.2% agarose gels in TAE buffer and stained post-electrophoresis using GelRed (Biotium, Inc.).

Supporting Information

S1 Fig. Alignment of predicted rainbowfish ERα transcript variants with the Japanese medaka *esr1* gene sequence.
(PDF)

S1 Table. Data yields for six RNA-Seq libraries generated from male and female rainbowfish tissues and sequenced using two lanes of an Illumina HiSeq 2000 instrument.
(DOCX)

S2 Table. Putative rainbowfish transcripts and coding sequence identifiers homologous to core eukaryotic genes (CEGs) from the genome of Japanese medaka (*Oryzias latipes*).
(CSV)

S3 Table. Putative rainbowfish transcripts and coding sequence identifiers homologous to the entire predicted from the genome of Japanese medaka (*Oryzias latipes*).
(CSV)

S4 Table. Rainbowfish transcripts encoding candidate nuclear receptors identified using conserved domain searches. Transcripts listed contain putative coding sequences matching conserved DNA-binding domains (Pfam cl02596; 'NR_DBD_like') and ligand-binding domains (Pfam cl11397; 'NR_LBD') from known nuclear receptors.
(CSV)

S5 Table. Rainbowfish transcripts encoding candidate nuclear receptors. Transcripts were identified by sequence similarity to predicted nuclear receptors from Japanese medaka (*Oryzias latipes*; obtained from KEGG BRITE ola03310).
(DOCX)

S6 Table. Rainbowfish transcripts encoding candidate steroidogenic enzymes. Transcripts were identified by sequence similarity to predicted steroidogenic enzymes from Japanese medaka (*Oryzias latipes*; obtained from KEGG pathway 00140).
(CSV)

S7 Table. Rainbowfish transcripts encoding candidate cytochrome P450-family enzymes. Transcripts were identified by sequence similarity with zebrafish CYP450 protein sequences from the Ensembl zebrafish genome assembly (Zv9) annotated as members of PFAM family PF00067 (cytochrome P450).
(CSV)

S8 Table. Rainbowfish transcripts encoding proteins with putative aryl hydrocarbon receptor activity. Transcripts were annotated via the JAMp workflow as being associated with Gene Ontology Molecular Function GO: 4874.
(CSV)

Acknowledgments

Author Contributions

Conceived and designed the experiments: PAB AK AP. Performed the experiments: PAB AP. Analyzed the data: AP PAB. Contributed reagents/materials/analysis tools: AK PAB AP. Wrote the paper: PAB. Designed the JAMp software used in the analysis: AP.

References

1. ABS. Gross Value of Irrigated Agricultural Production, 2011–12: Australian Bureau of Statistics; 2014 [2 Oct 2014].
2. Holdway DA, Barry MJ, Logan DC, Robertson D, Young V, Ahokas JT. Toxicity of pulse-exposed fenvalerate and esfenvalerate to larval australian crimson-spotted rainbow fish (*Melanotaenia fluviatilis*). Aquatic Toxicology. 1994; 28(3–4):169–87. doi: 10.1016/0166-445x(94)90032-9

3. Barry MJ, Ohalloran K, Logan DC, Ahokas JT, Holdway DA. Sublethal effects of esfenvalerate pulse-exposure on spawning and non-spawning australian crimson-spotted rainbowfish (*Melanotaenia fluviatilis*). Archives of Environmental Contamination and Toxicology. 1995; 28(4):459–63.

4. Barry MJ, Logan DC, Vandam RA, Ahokas JT, Holdway DA. Effect of age and weight-specific respiration rate on toxicity of esfenvalerate pulse-exposure to the australian crimson-spotted rainbow fish (*Melanotaenia fluviatilis*). Aquatic Toxicology. 1995; 32(2–3):115–26. doi: 10.1016/0166-445x(94)00084-4

5. Reid HP, Ahokas JT, Holdway DA. Use of cyanazine and malathion pulse-exposure toxicity to estimate the age-of-onset of functional liver-metabolism in larval australian crimson-spotted rainbowfish (*Melanotaenia fluviatilis*). Water Research. 1995; 29(8):2010–3. doi: 10.1016/0043-1354(95)00007-8

6. Barry MJ, Logan DC, Ahokas JT, Holdway DA. Effects of esfenvalerate poise-exposure on the survival and growth of larval australian crimson-spotted rainbow fish (*Melanotaenia fluviatilis*). Environmental Toxicology and Water Quality. 1995; 10(4):267–74. doi: 10.1002/tox.2530100406.

7. Kumar A, Chapman JC. Profenofos toxicity to the eastern rainbow fish (*Melanotaenia duboulayi*). Environmental Toxicology and Chemistry. 1998; 17(9):1799–806. doi: 10.1002/etc.5620170920

8. Williams ND, Holdway DA. The effects of pulse-exposed cadmium and zinc on embryo hatchability, larval development, and survival of Australian crimson spotted rainbow fish (*Melanotaenia fluviatilis*). Environmental Toxicology. 2000; 15(3):165–73.

9. Pollino CA, Holdway DA. Reproductive potential of crimson-spotted rainbowfish (*Melanotaenia fluviatilis*) following short-term exposure to bass strait crude oil and dispersed crude oil. Environmental Toxicology. 2002; 17(2):138–45. doi: 10.1002/tox.10043 PMID: 11979592

10. Pollino CA, Holdway DA. Toxicity testing of crude oil and related compounds using early life stages of the crimson-spotted rainbowfish (*Melanotaenia fluviatilis*). Ecotoxicology and Environmental Safety. 2002; 52(3):180–9. doi: 10.1006/eesa.2002.2190 PMID: 12297077

11. Vajda AM, Kumar A, Woods M, Williams M, Doan H, Tolsher P, et al. Integrated assessment of wastewater treatment plant effluent estrogenicity in the Upper Murray River, Australia, using the native Murray rainbowfish (*Melanotaenia fluviatilis*). Environmental Toxicology and Chemistry. 2015; 34(5):1078–87. doi: 10.1002/etc.2895 PMID: 25645549

12. Pollino CA, Georgiades E, Holdway DA. Use of the Australian crimson-spotted rainbowfish (*Melanotaenia fluviatilis*) as a model test species for investigating the effects of endocrine disruptors. Environmental Toxicology and Chemistry. 2007; 26(10):2171–8. doi: 10.1897/06-603r.1 PMID: 17867880

13. Holdway DA, Hefferman J, Smith A. Multigeneration assessment of nonylphenol and endosulfan using a model Australian freshwater fish, *Melanotaenia fluviatilis*. Environmental Toxicology. 2008; 23(2):253–62. doi: 10.1002/tox.20329 PMID: 18214909

14. Pollino CA, Georgiades E, Holdway DA. Physiological changes in reproductively active rainbowfish (*Melanotaenia fluviatilis*) following exposure to naphthalene. Ecotoxicology and Environmental Safety. 2009; 72(4):1265–70. doi: 10.1016/j.ecoenv.2009.01.012 PMID: 19286257

15. Shanthanagouda AH, Patil JG, Nugegoda D. Effects of exposure to oestrogenic compounds on aromatase gene expression are gender dependent in the rainbowfish, *Melanotaenia fluviatilis*. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology. 2013; 157(2):162–71. doi: 10.1016/j.cbpc.2012.11.004 PMID: 23178196

16. Woods M, Kumar A. Vitellogenin induction by 17β-estradiol and 17α-ethynylestradiol in male Murray rainbowfish (*Melanotaenia fluviatilis*). Environmental Toxicology and Chemistry. 2011; 30(11):2620–7. doi: 10.1002/etc.660 PMID: 21898560

17. Bhatia H, Kumar A, Du J, Chapman J, McLaughlin MJ. Di-n-butyl phthalate causes antiestrogenic effects in female murray rainbowfish (*Melanotaenia fluviatilis*). Environmental Toxicology and Chemistry. 2013; 32(10):2335–44. doi: 10.1002/etc.2304 PMID: 23761113

18. Bhatia H, Kumar A, Ogino Y, Gregg A, Chapman J, McLaughlin MJ, et al. Di-n-butyl phthalate causes estrogenic effects in adult male Murray rainbowfish (*Melanotaenia fluviatilis*). Aquatic Toxicology. 2014; 149:103–15. doi: 10.1016/j.aquatox.2014.01.025 PMID: 24576492

19. Bhatia H, Kumar A, Ogino Y, Du J, Gregg A, Chapman J, et al. Effects of the commercial antiandrogen flutamide on the biomarkers of reproduction in male murray rainbowfish (*Melanotaenia fluviatilis*). Environmental Toxicology and Chemistry. 2014; 33(5):1098–107. doi: 10.1002/etc.2524 PMID: 24453069

20. Jobling S, Beresford N, Nolan M, Rodgers-Gray T, Brighty GC, Sumpter JP, et al. Altered Sexual Maturation and Gamete Production in Wild Roach (*Rutilus rutilus*) Living in Rivers That Receive Treated Sewage Effluents. Biology of Reproduction. 2002; 66(2):272–81. doi: 10.1095/biolreprod66.2.272 PMID: 11804939

21. Kidd KA, Blanchfield PJ, Mills KH, Palace VP, Evans RE, Lazorchak JM, et al. Collapse of a fish population after exposure to a synthetic estrogen. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(21):8897–901. doi: 10.1073/pnas.0609568104 PMID: 17517636

22. Metcalfe CD, Metcalfe TL, Kiparissis Y, Koenig BG, Khan C, Hughes RJ, et al. Estrogenic potency of chemicals detected in sewage treatment plant effluents as determined by in vivo assays with Japanese medaka (*Oryzias latipes*). Environmental Toxicology and Chemistry. 2001; 20(2):297–308. PMID: 11351429

23. Jobling S, Nolan M, Tyler CR, Brighty G, Sumpter JP. Widespread Sexual Disruption in Wild Fish. Environmental Science & Technology. 1998; 32(17):2498–506. doi: 10.1021/es9710870

24. Katsu Y, Lange A, Urushitani H, Ichikawa R, Paull GC, Cahill LL, et al. Functional Associations between Two Estrogen Receptors, Environmental Estrogens, and Sexual Disruption in the Roach (*Rutilus rutilus*). Environmental Science & Technology. 2007; 41(9):3368–74. doi: 10.1021/es062797l

25. Lange A, Katsu Y, Miyagawa S, Ogino Y, Urushitani H, Kobayashi T, et al. Comparative responsiveness to natural and synthetic estrogens of fish species commonly used in the laboratory and field monitoring. Aquatic Toxicology. 2012; 109(0):250–8. doi: 10.1016/j.aquatox.2011.09.004

26. Miyagawa S, Lange A, Hirakawa I, Tohyama S, Ogino Y, Mizutani T, et al. Differing Species Responsiveness of Estrogenic Contaminants in Fish Is Conferred by the Ligand Binding Domain of the Estrogen Receptor. Environmental Science & Technology. 2014; 48(9):5254–63. doi: 10.1021/es5002659 PMID: 24689804

27. Pal A, Gin KY-H, Lin AY-C, Reinhard M. Impacts of emerging organic contaminants on freshwater resources: Review of recent occurrences, sources, fate and effects. Science of The Total Environment. 2010; 408(24):6062–9. doi: 10.1016/j.scitotenv.2010.09.026 PMID: 20934204

28. Kumar A, Williams M, Woods M, Kookana R, Barber L, Vajda A, et al. Treated effluent in the aquatic environment: impact assessment of endocrine disrupting chemicals. Report to the National Water Commission EP12292. CSIRO: Water for a Healthy Country National Research Flagship, 2012.

29. Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, et al. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 2004; 431 (7011):946–57. PMID: 15496914

30. Hoegg S, Brinkmann H, Taylor J, Meyer A. Phylogenetic Timing of the Fish-Specific Genome Duplication Correlates with the Diversification of Teleost Fish. J Mol Evol. 2004; 59(2):190–203. doi: 10.1007/s00239-004-2613-z PMID: 15486693

31. Amores A, Force A, Yan Y-L, Joly L, Amemiya C, Fritz A, et al. Zebrafish hox Clusters and Vertebrate Genome Evolution. Science. 1998; 282(5394):1711–4. doi: 10.1126/science.282.5394.1711 PMID: 9831563

32. Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan Y- L, Kelly PD, Chu F, et al. Zebrafish Comparative Genomics and the Origins of Vertebrate Chromosomes. Genome Research. 2000; 10(12):1890–902. doi: 10.1101/gr.164800 PMID: 11116085

33. Postlethwait J, Amores A, Cresko W, Singer A, Yan Y- L. Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends in Genetics. 2004; 20(10):481–90. doi: 10.1016/j.tig.2004.08.001 PMID: 15363902

34. Cosnefroy A, Brion F, Maillot-Maréchal E, Porcher J-M, Pakdel F, Balaguer P, et al. Selective Activation of Zebrafish Estrogen Receptor Subtypes by Chemicals by Using Stable Reporter Gene Assay Developed in a Zebrafish Liver Cell Line. Toxicological Sciences. 2012; 125(2):439–49. doi: 10.1093/toxsci/kfr297 PMID: 22045033

35. Legler J, Zeinstra LM, Schuitemaker F, Lanser PH, Bogerd J, Brouwer A, et al. Comparison of in Vivo and in Vitro Reporter Gene Assays for Short-Term Screening of Estrogenic Activity. Environmental Science & Technology. 2002; 36(20):4410–5. doi: 10.1021/es010323a PMID: 12387416

36. Ogino Y, Katoh H, Kuraku S, Yamada G. Evolutionary History and Functional Characterization of Androgen Receptor Genes in Jawed Vertebrates. Endocrinology. 2009; 150(12):5415–27. doi: 10.1210/en.2009-0523 PMID: 19819965

37. Bain PA, Ogino Y, Miyagawa S, Iguchi T, Kumar A. Differential ligand selectivity of androgen receptors α and β from Murray–Darling rainbowfish (*Melanotaenia fluviatilis*). General and Comparative Endocrinology. 2015; 212(0):84–91. doi: 10.1016/j.ygcen.2015.01.024

38. Douard V, Brunet F, Boussau B, Ahrens-Fath I, Vlaeminck-Guillem V, Haendler B, et al. The fate of the duplicated androgen receptor in fishes: a late neofunctionalization event? BMC Evolutionary Biology. 2008; 8(1):336.

39. Guyomard R, Boussaha M, Krieg F, Hervet C, Quillet E. A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. BMC Genetics. 2012; 13(1):15.

40. Hook SE, Osborn HL, Spadaro DA, Simpson SL. Assessing mechanisms of toxicant response in the amphipod *Melita plumulosa* through transcriptomic profiling. Aquatic Toxicology. 2014; 146(0):247–57. doi: 10.1016/j.aquatox.2013.11.001

41. Hook SE, Osborn HL, Golding LA, Spadaro DA, Simpson SL. Dissolved and Particulate Copper Exposure Induces Differing Gene Expression Profiles and Mechanisms of Toxicity in the Deposit Feeding Amphipod Melita plumulosa. Environmental Science & Technology. 2014; 48(6):3504–12. doi: 10.1021/es405322s PMID: 24552435

42. Liu N-Y, Xu W, Papanicolaou A, Dong S-L, Anderson A. Identification and characterization of three chemosensory receptor families in the cotton bollworm Helicoverpa armigera. BMC Genomics. 2014; 15(1):597.

43. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protocols. 2013; 8(8):1494–512. doi: 10.1038/nprot.2013.084 PMID: 23845962

44. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech. 2011; 29(7):644–52.

45. Bairoch A, Boeckmann B, Ferro S, Gasteiger E. Swiss-Prot: Juggling between evolution and stability. Briefings in Bioinformatics. 2004; 5(1):39–55. doi: 10.1093/bib/5.1.39 PMID: 15153305

46. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25(1):25–9. Epub 2000/05/10. doi: 10.1038/75556 PMID: 10802651

47. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014; 42(Database issue):D199–205. Epub 2013/11/12. doi: 10.1093/nar/gkt1076 PMID: 24214961

48. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, et al. eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Research. 2013.

49. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007; 23(9):1061–7. doi: 10.1093/bioinformatics/btm071 PMID: 17332020

50. Udy DB, Voorhies M, Chan PP, Lowe TM, Dumont S. Draft *De Novo* Transcriptome of the Rat Kangaroo *Potorous tridactylus* as a Tool for Cell Biology. PLoS ONE. 2015; 10(8):e0134738. doi: 10.1371/journal.pone.0134738 PMID: 26252667

51. Zhang Z, Wang Y, Wang S, Liu J, Warren W, Mitreva M, et al. Transcriptome Analysis of Female and Male *Xiphophorus maculatus* Jp 163 A. PLoS ONE. 2011; 6(4):e18379. doi: 10.1371/journal.pone.0018379 PMID: 21483681

52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990; 215(3):403–10. Epub 1990/10/05. doi: 10.1016/s0022-2836(05)80360-2 PMID: 2231712

53. Kookana RS, Williams M, Boxall ABA, Larsson DGJ, Gaw S, Choi K, et al. Potential ecological footprints of active pharmaceutical ingredients: an examination of risk factors in low-, middle- and high-income countries. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2014; 369(1656). doi: 10.1098/rstb.2013.0586

54. Liu J-L, Wong M-H. Pharmaceuticals and personal care products (PPCPs): A review on environmental contamination in China. Environment International. 2013; 59(0):208–24. doi: 10.1016/j.envint.2013.06.012

55. Heberer T. Occurrence, fate, and removal of pharmaceutical residues in the aquatic environment: a review of recent research data. Toxicology Letters. 2002; 131(1–2):5–17. doi: 10.1016/S0378-4274(02)00041-3 PMID: 11988354

56. Yu Y, Wu L, Chang AC. Seasonal variation of endocrine disrupting compounds, pharmaceuticals and personal care products in wastewater treatment plants. Science of The Total Environment. 2013; 442(0):310–6. doi: 10.1016/j.scitotenv.2012.10.001

57. Martín J, Camacho-Muñoz D, Santos JL, Aparicio I, Alonso E. Occurrence of pharmaceutical compounds in wastewater and sludge from wastewater treatment plants: Removal and ecotoxicological impact of wastewater discharges and sludge disposal. Journal of Hazardous Materials. 2012; 239–240(0):40–7. doi: 10.1016/j.jhazmat.2012.04.068

58. Kasprzyk-Hordern B, Dinsdale RM, Guwy AJ. The removal of pharmaceuticals, personal care products, endocrine disruptors and illicit drugs during wastewater treatment and its impact on the quality of receiving waters. Water Research. 2009; 43(2):363–80. doi: 10.1016/j.watres.2008.10.047 PMID: 19022470

59. Halling-Sørensen B, Nors Nielsen S, Lanzky PF, Ingerslev F, Holten Lützhøft HC, Jørgensen SE. Occurrence, fate and effects of pharmaceutical substances in the environment- A review. Chemosphere. 1998; 36(2):357–93. doi: 10.1016/S0045-6535(97)00354-8 PMID: 9569937

60. Kolpin DW, Furlong ET, Meyer MT, Thurman EM, Zaugg SD, Barber LB, et al. Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999–2000: A national reconnaissance. Environmental Science & Technology. 2002; 36(6):1202–11. PMID: 11944670

61. Fent K, Weston AA, Caminada D. Ecotoxicology of human pharmaceuticals. Aquatic Toxicology. 2006; 76(2):122–59. doi: 10.1016/j.aquatox.2005.09.009 PMID: 16257063

62. LaLone CA, Villeneuve DL, Olmstead AW, Medlock EK, Kahl MD, Jensen KM, et al. Effects of a glucocorticoid receptor agonist, dexamethasone, on fathead minnow reproduction, growth, and development. Environmental Toxicology and Chemistry. 2012; 31(3):611–22. doi: 10.1002/etc.1729 PMID: 22189798

63. Zeilinger J, Steger-Hartmann T, Maser E, Goller S, Vonk R, Länge R. Effects of synthetic gestagens on fish reproduction. Environmental Toxicology and Chemistry. 2009; 28(12):2663–70. doi: 10.1897/08-485.1 PMID: 19469587

64. Paulos P, Runnalls TJ, Nallani G, La Point T, Scott AP, Sumpter JP, et al. Reproductive responses in fathead minnow and Japanese medaka following exposure to a synthetic progestin, Norethindrone. Aquatic Toxicology. 2010; 99(2):256–62. doi: 10.1016/j.aquatox.2010.05.001 PMID: 20617545

65. Svensson J, Fick J, Brandt I, Brunström B. Environmental concentrations of an androgenic progestin disrupts the seasonal breeding cycle in male three-spined stickleback (*Gasterosteus aculeatus*). Aquatic Toxicology. 2014; 147(0):84–91. doi: 10.1016/j.aquatox.2013.12.013

66. Ellestad LE, Cardon M, Chambers IG, Farmer JL, Hartig P, Stevens K, et al. Environmental Gestagens Activate Fathead Minnow (*Pimephales promelas*) Nuclear Progesterone and Androgen Receptors in Vitro. Environmental Science & Technology. 2014; 48(14):8179–87. doi: 10.1021/es501428u PMID: 24911891

67. Bain PA, Kumar A, Ogino Y, Iguchi T. Nortestosterone-derived synthetic progestogens do not activate the progestogen receptor of Murray–Darling rainbowfish (*Melanotaenia fluviatilis*) but are potent agonists of androgen receptors alpha and beta. Aquatic Toxicology. 2015; 163(0):97–101. doi: 10.1016/j.aquatox.2015.03.021

68. Liu G, Moon TW, Metcalfe CD, Lee LE, Trudeau VL. A teleost in vitro reporter gene assay to screen for agonists of the peroxisome proliferator-activated receptors. Environmental toxicology and chemistry / SETAC. 2005; 24(9):2260–6. Epub 2005/10/01. PMID: 16193754

69. Leaver MJ, Boukouvala E, Antonopoulou E, Diez A, Favre-Krey L, Ezaz MT, et al. Three Peroxisome Proliferator-Activated Receptor Isotypes from Each of Two Species of Marine Fish. Endocrinology. 2005; 146(7):3150–62. doi: 10.1210/en.2004-1638 PMID: 15790725

70. Robinson-Rechavi M, Garcia HE, Laudet V. The nuclear receptor superfamily. Journal of Cell Science. 2003; 116(4):585–6. doi: 10.1242/jcs.00247

71. Germain P, Staels B, Dacquet C, Spedding M, Laudet V. Overview of Nomenclature of Nuclear Receptors. Pharmacological Reviews. 2006; 58(4):685–704. doi: 10.1124/pr.58.4.2 PMID: 17132848

72. Svensson J, Fick J, Brandt I, Brunström B. The Synthetic Progestin Levonorgestrel Is a Potent Androgen in the Three-Spined Stickleback (*Gasterosteus aculeatus*). Environmental Science & Technology. 2013; 47(4):2043–51. doi: 10.1021/es304305k PMID: 23362984

73. Overturf MD, Overturf CL, Carty DR, Hala D, Huggett DB. Levonorgestrel exposure to fathead minnows (*Pimephales promelas*) alters survival, growth, steroidogenic gene expression and hormone production. Aquatic Toxicology. 2014; 148(0):152–61. doi: 10.1016/j.aquatox.2014.01.012

74. Whitehead SA, Rice S. Endocrine-disrupting chemicals as modulators of sex steroid synthesis. Best Practice & Research Clinical Endocrinology & Metabolism. 2006; 20(1):45–61. doi: 10.1016/j.beem.2005.09.003 PMID: 16522519

75. Sanderson JT, Boerma J, Lansbergen GWA, van den Berg M. Induction and Inhibition of Aromatase (CYP19) Activity by Various Classes of Pesticides in H295R Human Adrenocortical Carcinoma Cells. Toxicology and Applied Pharmacology. 2002; 182(1):44–54. doi: 10.1006/taap.2002.9420 PMID: 12127262

76. Ankley GT, Kahl MD, Jensen KM, Hornung MW, Korte JJ, Makynen EA, et al. Evaluation of the Aromatase Inhibitor Fadrozole in a Short-Term Reproduction Assay with the Fathead Minnow (*Pimephales promelas*). Toxicological Sciences. 2002; 67(1):121–30. doi: 10.1093/toxsci/67.1.121 PMID: 11961225

77. Piferrer F, Zanuy S, Carrillo M, Solar II, Devlin RH, Donaldson EM. Brief treatment with an aromatase inhibitor during sex differentiation causes chromosomally female salmon to develop as normal, functional males. Journal of Experimental Zoology. 1994; 270(3):255–62. doi: 10.1002/jez.1402700304

78. Callard GV, Tchoudakova AV, Kishida M, Wood E. Differential tissue distribution, developmental programming, estrogen regulation and promoter characteristics of cyp19 genes in teleost fish. The Journal of Steroid Biochemistry and Molecular Biology. 2001; 79(1–5):305–14. doi: 10.1016/S0960-0760 (01)00147-9 PMID: 11850237

79. Halm S, Pounds N, Maddix S, Rand-Weaver M, Sumpter JP, Hutchinson TH, et al. Exposure to exogenous 17β-oestradiol disrupts P450aromB mRNA expression in the brain and gonad of adult fathead minnows (Pimephales promelas). Aquatic Toxicology. 2002; 60(3–4):285–99. doi: 10.1016/S0166-445X(02)00011-5 PMID: 12200092

80. Sanderson JT. The Steroid Hormone Biosynthesis Pathway as a Target for Endocrine-Disrupting Chemicals. Toxicological Sciences. 2006; 94(1):3–21. doi: 10.1093/toxsci/kfl051 PMID: 16807284

81. Noaksson E, Linderoth M, Bosveld ATC, Balk L. Altered steroid metabolism in several teleost species exposed to endocrine disrupting substances in refuse dump leachate. General and Comparative Endocrinology. 2003; 134(3):273–84. doi: 10.1016/S0016-6480(03)00267-3 PMID: 14636634

82. Ankley GT, Jensen KM, Kahl MD, Makynen EA, Blake LS, Greene KJ, et al. Ketoconazole in the fathead minnow (Pimephales promelas): Reproductive toxicity and biological compensation. Environmental Toxicology and Chemistry. 2007; 26(6):1214–23. doi: 10.1897/06-428r.1 PMID: 17571688

83. Gelinas D, Pitoc GA, Callard GV. Isolation of a goldfish brain cytochrome P450 aromatase cDNA:: mRNA expression during the seasonal cycle and after steroid treatment. Molecular and Cellular Endocrinology. 1998; 138(1–2):81–93. doi: 10.1016/S0303-7207(98)00015-X PMID: 9685217

84. Villeneuve DL, Blake LS, Brodin JD, Greene KJ, Knoebl I, Miracle AL, et al. Transcription of Key Genes Regulating Gonadal Steroidogenesis in Control and Ketoconazole- or Vinclozolin-Exposed Fathead Minnows. Toxicological Sciences. 2007; 98(2):395–407. doi: 10.1093/toxsci/kfm124 PMID: 17517826

85. Häggström M, Richfield D. Diagram of the pathways of human steroidogenesis. Wikiversity Journal of Medicine. 2014. doi: 10.15347/wjm/2014.005

86. Leitz T, Reinboth R. The biosynthesis of 11-ketotestosterone by the testis of the Siamese fighting fish Betta splendens regan (anabantoidei, belontiidae). General and Comparative Endocrinology. 1987; 66(1):145–57. doi: 10.1016/0016-6480(87)90359-5 PMID: 3582944

87. Martyniuk CJ, Bissegger S, Langlois VS. Current perspectives on the androgen 5 alpha-dihydrotestosterone (DHT) and 5 alpha-reductases in teleost fishes and amphibians. General and Comparative Endocrinology. 2013; 194(0):264–74. doi: 10.1016/j.ygcen.2013.09.019

88. Diotel N, Do-Rego J- L, Anglade I, Vaillant C, Pellegrini E, Vaudry H, et al. The brain of teleost fish, a source and a target of sexual steroids. Frontiers in Neuroscience. 2011; 5. doi: 10.3389/fnins.2011.00137 PMID: 22194715

89. Margiotta-Casaluci L, Courant F, Antignac J-P, Le Bizec B, Sumpter JP. Identification and quantification of 5α-dihydrotestosterone in the teleost fathead minnow (Pimephales promelas) by gas chromatography–tandem mass spectrometry. General and Comparative Endocrinology. 2013; 191(0):202–9. doi: 10.1016/j.ygcen.2013.06.017

90. Sumpter JP, Jobling S. Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. Environ Health Perspect. 1995; 103 Suppl 7:173–8. Epub 1995/10/01. PMID: 8593867

91. Arukwe A, Knudsen FR, Goksoyr A. Fish zona radiata (eggshell) protein: a sensitive biomarker for environmental estrogens. Environ Health Perspect. 1997; 105(4):418–22. Epub 1997/04/01. PMID: 9189707

92. Denslow N, Chow M, Kroll K, Green L. Vitellogenin as a Biomarker of Exposure for Estrogen or Estrogen Mimics. Ecotoxicology. 1999; 8(5):385–98. doi: 10.1023/a:1008986522208

93. Filby AL, Tyler CR. Molecular Characterization of Estrogen Receptors 1, 2a, and 2b and Their Tissue and Ontogenic Expression Profiles in Fathead Minnow (Pimephales promelas). Biology of Reproduction. 2005; 73(4):648–62. doi: 10.1095/biolreprod.105.039701 PMID: 15930325

94. Sabo-Attwood T, Kroll KJ, Denslow ND. Differential expression of largemouth bass (Micropterus salmoides) estrogen receptor isotypes alpha, beta, and gamma by estradiol. Molecular and Cellular Endocrinology. 2004; 218(1–2):107–18. doi: 10.1016/j.mce.2003.12.007 PMID: 15130515

95. Yamaguchi A, Ishibashi H, Kohra S, Arizono K, Tominaga N. Short-term effects of endocrine-disrupting chemicals on the expression of estrogen-responsive genes in male medaka (Oryzias latipes). Aquatic Toxicology. 2005; 72(3):239–49. doi: 10.1016/j.aquatox.2004.12.011 PMID: 15820104

96. Black DL. Mechanisms of alternative pre-messenger rna splicing. Annual Review of Biochemistry. 2003; 72(1):291–336. doi: 10.1146/annurev.biochem.72.121801.161720

97. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008; 40(12):1413–5. doi: 10.1038/ng.259 PMID: 18978789

98. Cotter KA, Nacci D, Champlin D, Chuprin J, Callard GV. Cloning of multiple ERα mRNA variants in killifish (*Fundulus heteroclitus*), and differential expression by tissue type, stage of reproduction, and estrogen exposure in fish from polluted and unpolluted environments. Aquatic Toxicology. 2015; 159 (0):184–97. http://dx.doi.org/10.1016/j.aquatox.2014.12.012

99. Cotter KA, Yershov A, Novillo A, Callard GV. Multiple structurally distinct ERα mRNA variants in zebrafish are differentially expressed by tissue type, stage of development and estrogen exposure. General and Comparative Endocrinology. 2013; 194(0):217–29. doi: 10.1016/j.ygcen.2013.09.014

100. Pakdel F, Métivier R, Flouriot G, Valotaire Y. Two Estrogen Receptor (ER) Isoforms with Different Estrogen Dependencies Are Generated from the Trout ER Gene. Endocrinology. 2000; 141(2):571–80. doi: 10.1210/endo.141.2.7296 PMID: 10650938

101. Xia Z, Patiño R, Gale WL, Maule AG, Densmore LD. Cloning,in Vitro Expression, and Novel Phylogenetic Classification of a Channel Catfish Estrogen Receptor. General and Comparative Endocrinology. 1999; 113(3):360–8. doi: 10.1006/gcen.1999.7196 PMID: 10068497

102. Tan NS, Lam TJ, Ding JL. The first contiguous estrogen receptor gene from a fish, Oreochromis aureus: evidence for multiple transcripts. Molecular and Cellular Endocrinology. 1996; 120(2):177–92. doi: 10.1016/0303-7207(96)03836-1 PMID: 8832578

103. Seo JS, Lee Y-M, Jung S-O, Kim I-C, Yoon Y-D, Lee J-S. Nonylphenol modulates expression of androgen receptor and estrogen receptor genes differently in gender types of the hermaphroditic fish Rivulus marmoratus. Biochemical and Biophysical Research Communications. 2006; 346(1):213–23. doi: 10.1016/j.bbrc.2006.05.123 PMID: 16759644

104. Pinto PIS, Teodósio R, Socorro S, Power DM, Canário AVM. Structure, tissue distribution and estrogen regulation of splice variants of the sea bream estrogen receptor α gene. Gene. 2012; 503(1):18–24. doi: 10.1016/j.gene.2012.04.081 PMID: 22579469

105. Patiño R, Xia Z, Gale WL, Wu C, Maule AG, Chang X. Novel Transcripts of the Estrogen Receptor α Gene in Channel Catfish. General and Comparative Endocrinology. 2000; 120(3):314–25. doi: 10.1006/gcen.2000.7566 PMID: 11121296

106. Petit FG, Métivier R, Valotaire Y, Pakdel F. Synergism between a half-site and an imperfect estrogen-responsive element, and cooperation with COUP-TFI are required for estrogen receptor (ER) to achieve a maximal estrogen-stimulation of rainbow trout ER gene. European Journal of Biochemistry. 1999; 259(1–2):385–95. doi: 10.1046/j.1432-1327.1999.00072.x PMID: 9914518

107. Zhao Q-Y, Wang Y, Kong Y-M, Luo D, Li X, Hao P. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics. 2011; 12(Suppl 14):S2. doi: 10.1186/1471-2105-12-S14-S2 PMID: 22373417

108. Xu D-L, Long H, Liang J-J, Zhang J, Chen X, Li J-L, et al. De novo assembly and characterization of the root transcriptome of Aegilops variabilis during an interaction with the cereal cyst nematode. BMC Genomics. 2012; 13(1):133.

109. Duan J, Xia C, Zhao G, Jia J, Kong X. Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. BMC Genomics. 2012; 13(1):392.

110. Burset M, Seledtsov IA, Solovyev VV. Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Research. 2000; 28(21):4364–75. doi: 10.1093/nar/28.21.4364 PMID: 11058137

111. Urushitani H, Nakai M, Inanaga H, Shimohigashi Y, Shimizu A, Katsu Y, et al. Cloning and characterization of estrogen receptor α in mummichog, Fundulus heteroclitus. Molecular and Cellular Endocrinology. 2003; 203(1–2):41–50. doi: 10.1016/S0303-7207(03)00118-7 PMID: 12782402

112. Greytak SR, Callard GV. Cloning of three estrogen receptors (ER) from killifish (Fundulus heteroclitus): Differences in populations from polluted and reference environments. General and Comparative Endocrinology. 2007; 150(1):174–88. doi: 10.1016/j.ygcen.2006.07.017 PMID: 16962596

113. Nelson DR, Zeldin DC, Hoffman SM, Maltais LJ, Wain HM, Nebert DW. Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. Pharmacogenetics. 2004; 14(1):1–18. Epub 2004/05/07. PMID: 15128046

114. Bucheli TD, Fent K. Induction of cytochrome P450 as a biomarker for environmental contamination in aquatic ecosystems. Critical Reviews in Environmental Science and Technology. 1995; 25(3):201–68. doi: 10.1080/10643389509388479

115. Goldstone J, McArthur A, Kubota A, Zanette J, Parente T, Jonsson M, et al. Identification and developmental expression of the full complement of Cytochrome P450 genes in Zebrafish. BMC Genomics. 2010; 11(1):643.

116. Nelson DR. Comparison of P450s from human and fugu: 420 million years of vertebrate P450 evolution. Archives of Biochemistry and Biophysics. 2003; 409(1):18–24. doi: 10.1016/S0003-9861(02)00553-2 PMID: 12464240

117. Fukunaga BN, Probst MR, Reisz-Porszasz S, Hankinson O. Identification of Functional Domains of the Aryl Hydrocarbon Receptor. Journal of Biological Chemistry. 1995; 270(49):29270–8. doi: 10.1074/jbc.270.49.29270 PMID: 7493958

118. Hankinson O. The Aryl Hydrocarbon Receptor Complex. Annual Review of Pharmacology and Toxicology. 1995; 35(1):307–40. doi: 10.1146/annurev.pa.35.040195.001515

119. Schmidt JV, Bradfield CA. AH Receptor Signaling Pathways. Annual Review of Cell and Developmental Biology. 1996; 12(1):55–89. doi: 10.1146/annurev.cellbio.12.1.55

120. Andreasen EA, Hahn ME, Heideman W, Peterson RE, Tanguay RL. The Zebrafish (Danio rerio) Aryl Hydrocarbon Receptor Type 1 Is a Novel Vertebrate Receptor. Molecular Pharmacology. 2002; 62 (2):234–49. doi: 10.1124/mol.62.2.234 PMID: 12130674

121. Karchner SI, Powell WH, Hahn ME. Identification and Functional Characterization of Two Highly Divergent Aryl Hydrocarbon Receptors (AHR1 and AHR2) in the Teleost Fundulus heteroclitus: EVIDENCE FOR A NOVEL SUBFAMILY OF LIGAND-BINDING BASIC HELIX LOOP HELIX-PER-ARNT-SIM (bHLH-PAS) FACTORS. Journal of Biological Chemistry. 1999; 274(47):33814–24. doi: 10.1074/jbc.274.47.33814 PMID: 10559277

122. Hansson MC, Wittzell H, Persson K, von Schantz T. Characterization of two distinct aryl hydrocarbon receptor (AhR2) genes in Atlantic salmon (Salmo salar) and evidence for multiple AhR2 gene lineages in salmonid fish. Gene. 2003; 303(0):197–206. doi: 10.1016/S0378-1119(02)01178-2

123. Karchner SI, Franks DG, Hahn ME. AHR1B, a new functional aryl hydrocarbon receptor in zebrafish: tandem arrangement of ahr1b and ahr2 genes. The Biochemical journal. 2005; 392(Pt 1):153–61. Epub 2005/07/27. doi: 10.1042/bj20050713 PMID: 16042621

124. Li Z-H, Xu H, Zheng W, Lam SH, Gong Z. RNA-Sequencing Analysis of TCDD-Induced Responses in Zebrafish Liver Reveals High Relatedness to In Vivo Mammalian Models and Conserved Biological Pathways. PLoS ONE. 2013; 8(10):e77292. doi: 10.1371/journal.pone.0077292 PMID: 24204792

125. Watson JD, Prokopec SD, Smith AB, Okey AB, Pohjanvirta R, Boutros PC. TCDD dysregulation of 13 AHR-target genes in rat liver. Toxicology and Applied Pharmacology. 2014; 274(3):445–54. doi: 10.1016/j.taap.2013.12.004 PMID: 24355419

126. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics. 2014.

127. Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Research. 2010; 38(12):e131. doi: 10.1093/nar/gkq224 PMID: 20395217

128. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012; 28(23):3150–2. doi: 10.1093/bioinformatics/bts565 PMID: 23060610

129. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006; 22(13):1658–9. doi: 10.1093/bioinformatics/btl158 PMID: 16731699

130. Remmert M, Biegert A, Hauser A, Soding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Meth. 2012; 9(2):173–5.

131. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009; 10(1):421.

132. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005; 21 (18):3674–6. doi: 10.1093/bioinformatics/bti610 PMID: 16081474

133. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research. 2011; 39(suppl 1):D225–D9. doi: 10.1093/nar/gkq1189

134. R_Core_Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.