Supplemental Information

Genomic and phenotypic response of honyhead turbot exposed to municipal wastewater effluents

Doris E. Vidal-Dorsch, Steven M. Bay, Cataldo Ribecco, L. James Sprague, Mila Angert, Colleen Ludka, Oliana Carnevali, Darrin J. Greenstein, Daniel Schlenk, Kevin M. Kelley, Jesus A. Reyes, Shane Snyder, Brett Vanderford, Lan C. Wiborg, Dawn Petschauer, Roman Sasik, Michael Baker and Gary Hardiman

Microarray data analysis

The improved multispecies array contained an expanded probe content including targets with defined roles in endocrine pathways and processes (Baker, 2005, 2011; Blanchette et al., 2007; Filby et al., 2007; Ollikainen et al., 2006), in addition to well-defined biomarkers for contaminant exposure (Hutchinson et al., 2006; Katchamart et al., 2002). Statistical analysis of the microarray experiment involved two steps: normalization of microarray data as described previously using an mloess algorithm, and sorting of the genes according to interest (Baker et al., 2009). To investigate alterations in gene expression of controls and exposed fish, we used the Significance Analysis of Microarrays [SAM] procedure as described by Tusher et al. (Tusher et al., 2001) and its implementation in the official statistical package SAMR.

In order to not be unduly impressed by accidental small variances, we set the percentile of standard deviation values used for the exchangeability factor s0 in the test statistic to 90. We used a false discovery rate (FDR) threshold of 0.2 as the criterion of significance, except in the case of 0.5% SD effluent comparison to seawater, in which case we used FDR < 0.3. In the 0.5% LA effluent comparison to seawater, we do not estimate FDR at all; rather, we set a cutoff for the test statistic itself based on our experience with the other comparisons. Special treatment of these two cases was necessitated by a combination of smaller biological effect and a smaller group size (n = 1 in the 0.5% LA effluent group). SAM helped identify genes with statistically significant changes in expression by assigning a score on the basis of its change in gene expression relative to the standard deviation of repeated measurements for that gene.

Genes with scores greater than the predetermined threshold were considered to be significant. Scatter plots of the signal intensity for each microarray spot were log2 transformed

for comparison. Thus, the log2 signal for each gene for the sea water control was compared with that of signals from the corresponding gene in fish exposed to ammonia, 0.5% effluent, 5% effluent and 30 nM E2 for two weeks.

The FDR is a q-value (a measure of false positives) rather than a p-value (a measure of statistical significance). It should be emphasized that an FDR of 0.2 is very relevant when dealing with a relatively small number of differentially expressed transcripts, as in the case of the array data analyzed here. If there are 20 differentially expressed transcripts, with a q-value of 0.2, one would expect two false positives. If we were dealing with a larger gene set, say 2000 genes, and then a q-value of 0.2 would be unacceptable. However, based on the small number of differentially expressed transcripts being analyzed, a q-value of 0.2 is acceptable, (given that key findings are validated by qPCR).

Microarray Probe Validation

In a previous study by members of this group (Rouse et al., 2008, BMC Res Notes. 2008 Jul 11; 1:45) we examined the effects of probe variability, signal dynamic range and feature morphology and from this prior work have an excellent sense of the sensitivity of the array used in this study. We evaluated multiple probes for the same gene on this 15K array, using diagnostic heat maps to ensure probe validity and uniformity of performance before the data was considered in our downstream analysis. Data were averaged across replicate probes. As averaging is additive in nature, the non-detection of low abundance transcripts due to a dilution effect did not pose a problem. The replicate features can be considered as one giant or large probe and this additive effect provides advantages for low level transcripts, enabling their

detection. The microarray is based on the Agilent *in situ* synthesized oligonucleotide microarray, which is a highly sensitive platform (Hardiman, 2004).

Amplification and sequencing of hornyhead turbot mRNAs

We amplified partial turbot transcripts using conserved sequences from other fish species to guide the choice of primer design. Gene-specific primers were designed using Primer3 software (Rozen and Skaletsky, 2000). A 117 bp sequence was amplified from the *VTG1* transcript using the following primers; forward 5'- ATGAAGGGACAGACCTGTGG and reverse 5'- AACCCAGGAATGAGCATAGC. A 127 bp sequence was amplified from *VTG2* transcript using the following primers; forward 5'- ACTGGATGAGAGGCCAGACTT and reverse 5'- GGTAGAACCCAGGAATGAGC.

A 92 bp sequence was amplified from *CYP1A* transcript using the following primers; forward 5'- TCACTGTGAGGACAGGAAGCT and reverse 5'- CAGCACCGAACAGGTCATT. A 98 bp sequence was amplified from *β-actin* using the following primers; forward 5'- TCCCTGGAGAAGAGCTACGA and reverse 5'- AGGAAGGAAGGCTGGAAGAG.

For amplification and sequencing, oligonucleotide primers were obtained from Invitrogen. PCR was carried out in 50µl volume, 5µl (10X) reaction buffer 200mMTrisHCl/100mMKCl/100mM$(NH_4)_2SO_4$/20mM$MgSO_4$/1%TritonX/BSA/1mg/m, 0.1mM dNTPs, 2.5 units of *Pfu,* 1µl 1µM forward primers, 1µl 1µM reverse primers, 200ng turbot liver cDNA. After a denaturing step for 10 min at 96°C, touchdown amplification was performed with 35 cycles of 45s at 96°C, 45s at 55° to 50°C, in one degree increments and 1 min at 72°C.

All reactions were evaluated on a 1% agarose gel stained with ethidium bromide to validate the reaction and the products were directly sequenced using the respective forward and reverse PCR primers**.** All of the sequencing reads were subjected to a series of quality control measures, including a Phred quality score > 20, and manual trace inspection. The identity of each sequence was confirmed by performing BLAST searches of GenBank.

Quantitative real-time PCR analysis

Relative turbot mRNA transcript levels were measured by real-time quantitative (qPCR) in a LightCycler 480. Total RNA was extracted from hepatic turbot samples as described above and reverse-transcribed using the Roche Transcriptor kit and 50 ng cDNA were quantified using LightCycler 480 SYBR Green Master kit. For selection of housekeeping genes appropriate for normalizing the qPCR array data, we investigated three housekeeping genes, β-actin, G3PDH and 18srRNA. Of these three, G3PDH and β-actin performed equally well. The *β-actin* transcript was chosen to serve as an internal control for normalization.

Each sample was run in duplicate using the primer pairs described above and mean values were reported. Amplification and melting curve analyses for these assays indicated that the amplifications were specific. Normalized gene expression values were obtained using LightCycler Relative Quantification software. Relative gene copy numbers were derived using the formula $2^{\Delta CT}$ where ΔCT is the difference in amplification cycles required to detect amplification product from equal starting concentrations of turbot liver RNA (Untergasser et al., 2007).

REFERENCES

Baker, M.E., 2005. Xenobiotics and the evolution of multicellular animals: emergence and diversification of ligand-activated transcription factors. Integrative and Comparative Biology 45, 172-178.

Baker, M.E., 2011. Origin and diversification of steroids: co-evolution of enzymes and nuclear receptors. Mol Cell Endocrinol 334, 14-20.

Baker, M.E., Ruggeri, B., Sprague, L.J., Eckhardt-Ludka, C., Lapira, J., Wick, I., Soverchia, L., Ubaldi, M., Polzonetti-Magni, A.M., Vidal-Dorsch, D., Bay, S., Gully, J.R., Reyes, J.A., Kelley, K.M., Schlenk, D., Breen, E.C., Sasik, R., Hardiman, G., 2009. Analysis of Endocrine Disruption in Southern California Coastal Fish Using an Aquatic Multispecies Microarray. Environmental Health Perspectives 117, 223-230.

Blanchette, B., Feng, X., Singh, B.R., 2007. Marine glutathione S-transferases. Mar Biotechnol (NY) 9, 513-542.

Filby, A.L., Santos, E.M., Thorpe, K.L., Maack, G., Tyler, C.R., 2007. Gene expression profiling for understanding chemical causation of biological effects for complex mixtures: a case study on estrogens. Environ Sci Technol 41, 8187-8194.

Hardiman, G., 2004 Microarray platforms--comparisons and contrasts. Pharmacogenomics 5, 487-502.

Hutchinson, T.H., Ankley, G.T., Segner, H., Tyler, C.R., 2006. Screening and testing for endocrine disruption in fish-biomarkers as "signposts," not "traffic lights," in risk assessment. Environ Health Perspect 114 Suppl 1, 106-114.

Katchamart, S., Miranda, C.L., Henderson, M.C., Pereira, C.B., Buhler, D.R., 2002. Effect of xenoestrogen exposure on the expression of cytochrome P450 isoforms in rainbow trout liver. Environ Toxicol Chem 21, 2445-2451.

Ollikainen, N., Chandsawangbhuwana, C., Baker, M.E., 2006. Evolution of the thyroid hormone, retinoic acid, and liver X receptors. Integrative and Comparative Biology 46, 815-826.

Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers, in: Misener, S., A., K.S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press Inc, Totowa (NJ), pp. 365-386.

Tusher, V.G., Tibshirani, R., Chu, G., 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98, 5116-5121.

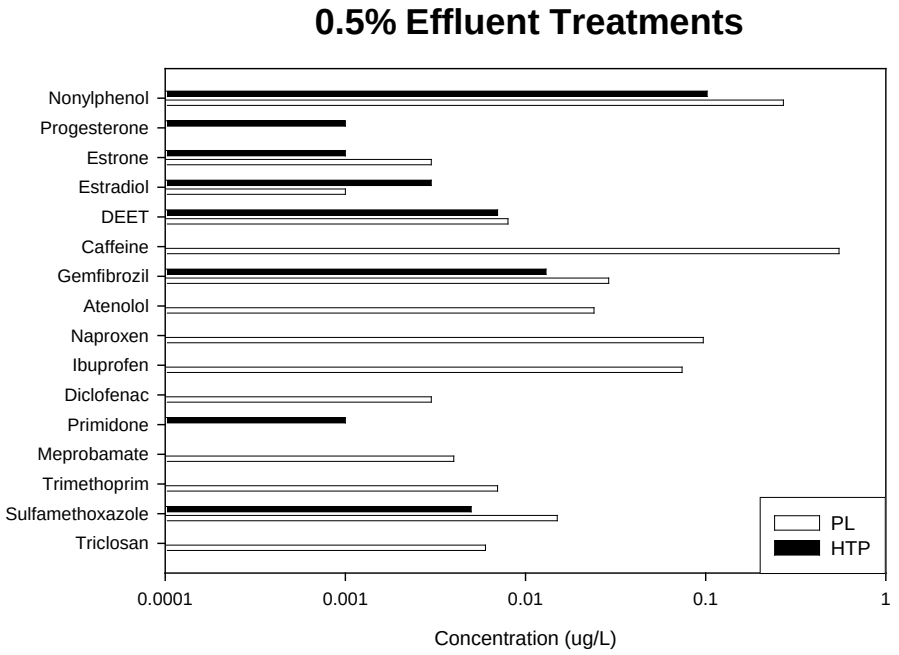
Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35, W71-74.

Table SI-1. Genes contained in the microarray (additional functions are not described).

Gene ID	Gene Name	Function
FXR	Farnesoid X receptor	Bile biosynthesis/transcription factor
HSP70	Heat shock protein 70	Cellular stress response
HSP90	Heat shock protein 90	Cellular stress response
AHR	Aryl hydrocarbon receptor	Contaminant response/Detoxification
CYP1A	Cytochrome P450, family 1, subfamily A	Contaminant response/Detoxification
CYP3A	Cytochrome P450, family 3, subfamily A	Contaminant response/Detoxification
GSTα	Glutathione S-transferase	Contaminant response/Detoxification
MT1	Metallothionein	Contaminant response/Detoxification
IGFBP1	Insulin-like growth factor building protein1	Development/Growth
IGFBP2	Insulin-like growth factor-binding protein 2	Development/Growth
IGFBP5	Insulin-like growth factor-binding protein 5	Development/Growth
VEGF	Vascular endothelial growth factor	Development/Growth
TRα	Thyroid receptor (alpha)	Development/mediates thyroid activities
TRβ	Thyroid receptor (beta)	Development/mediates thyroid activities
G3PDH	Glyceraldehyde 3-phosphate dehydrogenase	Glycolysis/nuclear functions
GRF1	Glucocorticoid receptor 1	Hormone signaling
HAMP	Hepcidin	Infection response
PPARɣ	Peroxisome proliferator activated receptor gamma	Lipid metabolism/development/transcription factor
PPARα	Peroxisome proliferator activated receptor alpha	Lipid metabolism/development/transcription factor
LXR	Liver X receptor	Lipid metabolism/transcription factor
ZP2	Zona pellucida glycoprotein 2	Reproduction/egg: sperm binding
ZP3	Zona pellucida glycoprotein 3	Reproduction/egg: sperm binding
VTG1	Vitellogenin A	Reproduction/egg: yolk protein
VTG2	Vitellogenin B	Reproduction/egg: yolk protein
CYP19	Cytochrome P450, family 19 (Aromatase)	Reproduction/Steroidogenesis
HSD17B1	Hydroxysteroid (17-beta) dehydrogenase type 1	Reproduction/Steroidogenesis
CYP11B	Cytochrome P450, family 11, subfamily B1	Reproduction/Steroidogenesis/stress
HSD11B1	Hydroxysteroid (11-beta) dehydrogenase 1	Reproduction/Steroidogenesis/stress
AR	Androgen receptor	Reproduction/transcription factor
ER1	Estrogen receptor alpha	Reproduction/transcription factor
ER2	Estrogen receptor beta	Reproduction/transcription factor
PR	Progesterone receptor	Reproduction/transcription factor
RAR	Retinoic acid receptor	Retinoid hormone transcription factor
RXR	Retinoid receptor	Retinoid hormone transcription factor
SHBG	Sex hormone-binding globulin	Sex steroid hormone binding
MR	Mineralocorticoid receptor	Steroid hormone transcription factor
VDR	Vitamin D receptor	Steroid hormone transcription factor
StAR	Steroidogenesis acute regulator	Steroidogenesis/other endocrine tissues

Table SI-2. List of analytes investigated and their reporting limits (RL), grouped as current-use pesticides (CUP), hormones, industrial and commercial compounds (ICC), and pharmaceuticals and personal care products (PPCP).

Group	Chemical Name	RL
CUP	Atrazine	0.0003
CUP	Lindane	0.001
Hormones	Estradiol	0.0005
Hormones	Estrone	0.0002
Hormones	Etynylestradiol	0.001
Hormones	Progesterone	0.0005
Hormones	Testosterone	0.0005
ICC	Benzophenone	0.053
ICC	Bisphenol A	0.05
ICC	Nonylphenol	0.08
ICC	Tris(1-chloro-2-propyl) phosphate (TCPP)	0.01
ICC	Tris(2-chloroethyl) phosphate (TCEP)	0.1
PPCP	Atenolol	0.001
PPCP	Atorvastatin	0.003
PPCP	Caffeine	0.005
PPCP	Carbamazepine	0.005
PPCP	DEET	0.001
PPCP	Diazepam	0.0003
PPCP	Diclofenac	0.003
PPCP	Dilantin	0.001
PPCP	Fluoxetine	0.0005
PPCP	Gemfibrozil	0.003
PPCP	Ibuprofen	0.05
PPCP	Iopromide	0.011
PPCP	Meprobamate	0.003
PPCP	Musk Ketone	0.025
PPCP	Naproxen	0.005
PPCP	Primidone	0.0005
PPCP	Sulfamethoxazole	0.003
PPCP	Triclosan	0.01
PPCP	Trimethoprim	0.003



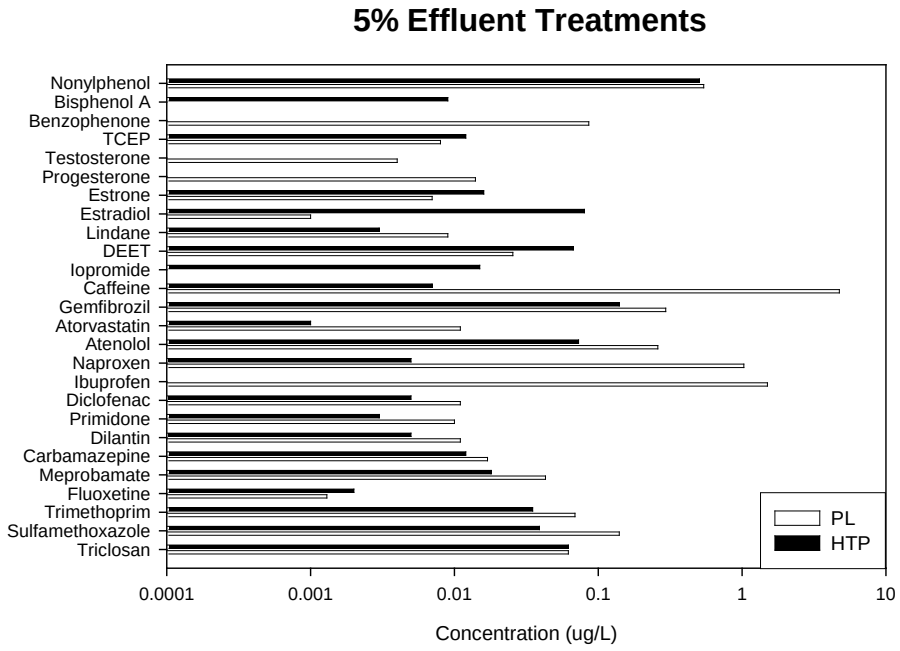


Figure SI-1. Average concentrations for compounds detected in advanced-primary treated effluent (PL) and full-secondary treated effluent (HTP) in exposure tanks.