Environmental Science & Technology

Article

Cite This: Environ. Sci. Technol. 2018, 52, 317−326

pubs.acs.org/est

Gene Coexpression Networks Drive and Predict Reproductive Effects in *Daphnia* in Response to Environmental Disturbances

J. Asselman,*[,†] M. E. Pfrender,[‡,§] J. A. Lopez,[§] J. R. Shaw,[∥,⊥] and K. A. C. De Schamphelaere[†]

[†]Laboratory of Environmental Toxicology and Aquatic Ecology, Environmental Toxicology Unit (GhEnToxLab), Ghent University, Ghent, B-9000, Belgium

[‡]Department of Biological Sciences and Environmental Change Initiative, University of Notre Dame, Indiana 46556, United States

[§]Genomics & Bioinformatics Core, University of Notre Dame, Indiana 46556, United States

[∥]The School of Public and Environmental Affairs and The Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana, United States

[⊥]Environmental Genomics Group, School of Biosciences, University of Birmingham, Birmingham, U.K.

Supporting Information

ABSTRACT: Increasing effects of anthropogenic stressors and those of natural origin on aquatic ecosystems have intensified the need for predictive and functional models of their effects. Here, we use gene expression patterns in combination with weighted gene coexpression networks and generalized additive models to predict effects on reproduction in the aquatic microcrustacean *Daphnia*. We developed models to predict effects on reproduction upon exposure to different cyanobacteria, different insecticides and binary mixtures of cyanobacteria and insecticides. Models developed specifically for groups of stressors (e.g., either cyanobacteria or insecticides) performed better than general models developed on all data. Furthermore, models developed using in silico generated mixture gene expression profiles from single stressor data were able to better predict effects on reproduction compared to models derived from the mixture exposures themselves. Our results highlight the potential of gene expression data to quantify effects of complex exposures at higher level organismal effects without prior mechanistic knowledge or complex exposure data.

1. INTRODUCTION

Increasing events of environmental perturbations in aquatic environments have increased the complexity of stressor exposures and intensified the need to understand and predict their effects on aquatic ecosystems.[1−3] These events, driven by both natural and anthropogenic-induced changes, can significantly affect species survival and reproduction, which are both key drivers of population growth.[4] Underlying these life-history parameters are coordinated interactions and changes in gene networks and pathways that trigger a cascade of responses at higher functional levels.[5] The first manifestations of environmental changes can therefore be detected at gene expression level, suggesting that gene expression patterns can be used as causal links between environmental disturbances and demographic parameters. Yet, despite the increasing amount of molecular data available, it remains a challenge to straightforwardly link gene expression patterns to higher organismal effects, for example, apical end points.[6]

More and more studies are developing predictive models to identify causal links between gene expression to other parameters of interest such as apical end points or the environment. Particularly in the area of human health, gene coexpression networks and gene expression profiles have been employed to identify causal links between genes and the onset of diseases such as cancer or apical end points such as weight.[7−13] In plant studies, gene network analyses have been used to link gene expression to different phenotypes, to phenotypic responses to environmental stress.[14,15] In environmental toxicology and aquatic ecology, these studies have primarily focused on linking gene expression to specific environmental stressors and have focused on using gene expression patterns to identify or classify the type of exposure.[16−19] However, little attention has been given to quantitatively link environmental stress on apical end points to gene expression patterns. Yet, quantitative models that link environmental stress to effects on apical end points via gene expression are necessary to predict effects and thoroughly assess potential impacts of current and future environmental changes such as increased chemical pollution or alterations in climate. These predictions become particularly challenging when considering the complex multistress real world environment.[20] Current risk assessments and environmental policies do not account for interactions between different chemicals, between chemicals and natural stressors.[20,21] As a result, their

Received: October 12, 2017
Revised: December 5, 2017
Accepted: December 6, 2017
Published: December 6, 2017

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326

predictive power and potential forecasting is limited and contrasts with the complexity of the actual environment.

In environmental toxicology and ecology, predictive models can provide better estimates of the potential effects of chemicals prior to their use, and the implications of their discharge in the environment. Current environmental policies are generally based on a small subset of standard toxicity testing with single chemicals, which contrasts with environmental reality where multiple stressors usually co-occur and affect organisms. Furthermore, current standard toxicity tests are time-consuming and limited to priority chemicals as they are exposure experiments in the laboratory during which effects are monitored. As such, they are not predictive, making it impossible to proactively take precautions that could mitigate potential future impacts, including the impact of climate change. In contrast, predictive models allow us to model and forecast a wide range of potential harmful and detrimental effects, including potential interactions between multiple stressors, and develop the best possible environmental management strategies to mitigate these effects. Here, we propose to develop generalized additive models in which gene expression networks can be used as predictors of reproductive effects in the aquatic crustacean *Daphnia* when exposed to different chemical and natural stressors, including their binary mixtures.

2. MATERIALS AND METHODS

2.1. Experimental Setup. Experimental organisms were harvested from parthenogenetic female laboratory cultures of *Daphnia pulex* maintained under identical conditions as the experiment. The isolates were obtained from the laboratory of Prof. Shaw (Indiana University, Bloomington, IN) and have been in culture in the current laboratory for more than 100 generations. Harvested animals were less than 24 h old and were assigned randomly to experimental treatments. Animals were cultured in COMBO medium without nitrogen and phosphorus stocks under a photoperiod of 16:8h light:dark in a climate control chamber at 20 ± 1 °C.[22] They were fed daily with 3 mg dry weight L^{-1} of an algal mixture consisting of *Pseudokirchneriella subcapitata* and *Chlamydomonas reinhardtii* in a 3:1 cell number ratio.[23] The experiment consisted of 48 binary mixtures between six different cyanobacteria from six different genera (Supporting Information (SI) Table S1) and eight insecticides that differed in their modes of action (SI Tables S2 and S3). We specifically selected combinations of cyanobacteria and insecticides as representatives of natural and chemical stressors specifically, as both comprise a diverse group of stressors with very different modes of actions (Table 1). In addition, our previous research has shown both synergistic and antagonistic interactions at life history level.[24,25] Also, given that eutrophication stimulates cyanobacterial blooms and that eutrophication commonly occurs in agricultural areas, it is cyanobacterial blooms and insecticides are likely to occur together in aquatic environments Each binary combination consisted of four treatments: control, insecticide, cyanobacteria and mixture treatment. All treatments consisted of five replicate beakers (1.5L), containing each 30 animals. On the fourth day, these animals were exposed for 10 days to their assigned treatment. In control conditions, no insecticide nor cyanobacteria suspension was added to the medium. In cyanobacteria treatments, the control diet was reduced with 50%, on dry weight basis and was contaminated with 50% of one of the six respective cyanobacteria. This diet ratio was selected based upon previous research as it induced clear gene expression patterns and life history changes.[26] Insecticide treatments contained a given amount of one of the eight insecticides (SI Table S3). The same effect concentration for reproduction was chosen for all insecticide treatments, that is, one-half of the EC50.[24,25] This effect concentration was selected as it was deemed high enough to elicit a toxic response, but low enough to allow quantification of potential synergisms in the mixture treatment. (i.e., if the concentration in the mixture treatment is too high, the effect approaches 100% which makes it impossible to quantify potential synergisms as they would be larger than 100%). Mixture treatments consisted of the same insecticide concentration as the insecticide treatment and animals were given the same diet as the cyanobacteria treatment.

Table 1. Overview of the Different Insecticides and Cyanobacteria Used in the Experiment[a]

insecticide (mode of action)	cyanobacteria
acetamiprid (agonist of nicotinic acetylcholine receptor)	*Anabaena lemmermannii*
carbaryl (inhibitor of acetylcholine esterase)	*Aphanizonmenon* sp.
chlorpyrifos (inhibitor of acetylcholine esterase)	*Cylindrospermopsis raciborskii*
deltamethrin (modulator of sodium channels)	*Microcystis aeruginosa*
endosulfan (antagonist of GABA-gated chloride channels)	*Nodularia* sp.
fenoxycarb (juvenile hormone mimic)	*Oscillatoria* sp.
tebufenpyrad (inhibitor of mitochondrial complex i electron transport)	*Anabaena lemmermannii*
tetradifon (inhibitor of mitochondrial ATP synthase)	

[a]Detailed information on the cyanobacteria strains can be found in SI Table S1.

For all treatments, medium was renewed every 2 days. At the same time, reproduction and survival were monitored. Reproduction was measured for the entire pool of animals and survival data was used to account for mortality when comparing reproduction across replicate beakers. If the animals reproduced, neonates were counted and removed from the beaker. At the end of the experiment, RNA was extracted from adult exposed animals for gene expression analysis at the end of the experiment. Samples for concentration analysis of insecticides were taken with every medium renewal of both old and new media and measured with gas chromatography after solid phase extractions, analysis protocols are described in detail in SI Tables S4–S6. At the same time, pH was measured for all treatments to ensure that pH never differed more than 0.2 units from control treatments (pH 7.00 ± 0.2).

2.2. Statistical Analysis. Significant differences between treatments were analyzed by comparing the total reproduction (number of neonates per beaker) relative to the total control reproduction across treatments in which each beaker was considered as an individual biological replicate. Analysis of variance with two factors was performed to determine interaction effects for each binary combination of cyanobacteria and insecticides on the log transformed total reproduction based on De Coninck et al.[27] Assumptions of normality and homoscedasticity were verified on the log transformed data with the Shapiro-Wilk test[28] and the Levene test,[29] respectively. All p-values were corrected for multiple testing applying the Benjamini-Hochberg false discovery rate (FDR) procedure at the 5% significance level.[30] Deviation from noninteraction and statistical significance were quantified according to De Coninck et al.[27] In particular, predicted reproduction is determined with

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326

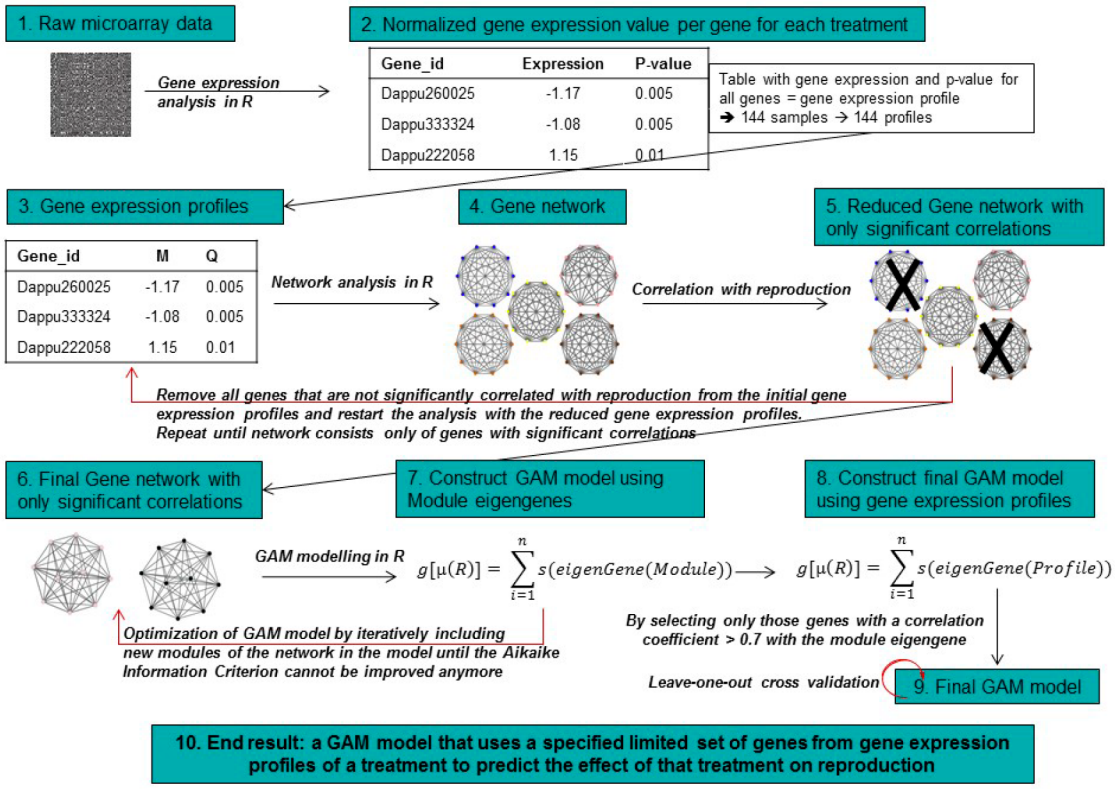

Figure 1. Schematic overview of the modeling approach using the package LIMMA (Linear Models for Microarray Data) and WGCNA(Weighted Gene Coexpression Network Analysis) within the R environment as well as generalized additive models (GAM).

the independent action model,[31] and thus estimated from the reproduction observed in the single stressor treatments. The absolute deviation is then quantified as the difference between the observed and predicted reproduction.

2.3. mRNA Extraction, Labeling, and Hybridization. RNA was extracted with the RNeasy kit and Qiashredder (Qiagen, Venlo, Netherlands) following manufacturer's protocol. All animals from one beaker were pooled into one sample and will further be referred to as one biological replicate. The microarray protocol can be found in the Supporting Information and was based on ref 32. The labeling design followed a standard loop design for each binary mixture to allow optimal comparison of single and combined effects within each mixture. Different biological replicates were used on each array for replication and dye swaps. All labeled samples were pooled according to the design, resulting in 336 pools to be hybridized across 32 12-plex arrays. The microarray itself is a transcriptome array developed by the Centre for Genomics and Bioinformatics (Indiana University, Bloomington, IN) and is in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus under the accession number (GEO: GPL11278). Arrays were scanned with the NimbleGen MS 200 Microarray Scanner to measure fluorescence and images were processed with NimbleScan 2.6 Software and deposited under the GEO accession number GSE102226.

2.4. Image Analysis and Data Processing. Microarray images were analyzed with the statistical software package R (R Development Core Team, 2011, version 3.0.1) and Bioconductor.[33] The LIMMA (,[34] version 3.16.7) package was used with additions and modifications according to Colbourne et al.[32] All signal distributions were quantile normalized across arrays, samples and replicates. Linear models were constructed with lmscFit function, which fits multiple linear models using least-squares and empirical Bayes Statistics were implemented with eBayes function.[34] For each treatment, a specific contrast analysis was then conducted to quantify the effect of that treatment relative to the control treatment (,[34] version 3.16.7). Benjamin-Hochberg method[30] was implemented to adjust p-values for multiple testing at a 95% significance level. Since the microarray development, the genome assembly of *Daphnia pulex* has improved significantly.[35] Therefore, we only used microarray probes that mapped uniquely to the improved genome assembly. We maintained the original identifiers that link probes to gene models as published by Colbourne et al.[32] All raw microarray data have been deposited in GEO (GSE102226).

2.5. Weighted Gene Coexpression Analysis (WGCNA). All data, both gene expression and reproduction responses, was integrated in a weighted gene coexpression analysis (,[36] WGCNA: version 1.27−1). The analysis was conducted in an iterative approach of network building by removing non-significant genes from the network, that is, genes that did not significantly correlate with changes in reproduction. The final networks were characterized by network modules, that is, groups of genes, and their eigengenes, which represent the overall expression profile of each module.[36]

2.6. Developing Generalized Additive Models to Predict Reproduction Using Gene Expression Data. We constructed generalized additive models (GAM) in which the eigengenes of the network modules were used as covariates to predict reproduction responses (Figure 1).[37] Therefore, we started with the module of which the eigengene had the highest correlation coefficient with the life-history variable, reproduction. We continued by subsequently adding other modules to

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326

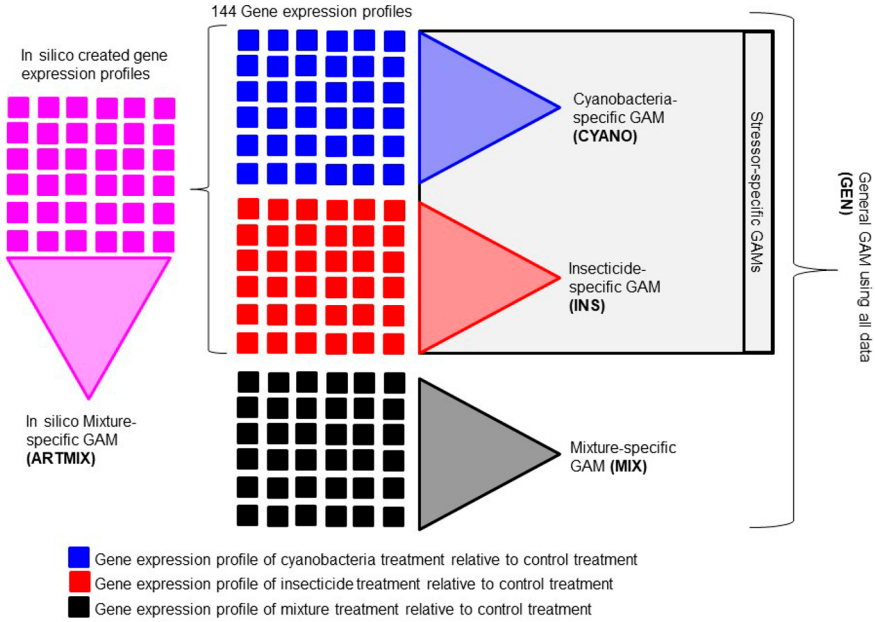

Figure 2. Schematic overview of the different generalized additive models (GAM) that will be constructed and the different data sets used for each model.

the model until the Aikaike Information Criterion (AIC) did not improve. Each time, the module that resulted in the best model improvement, by evaluating the AIC, was added to the model. Once the final model was selected, we reconstructed those network modules that served as input for the model, by only retaining genes in the network of which the expression profile showed a high correlation coefficient (0.75 or higher) with the module eigengene. We then used these smaller reconstructed networks in the final model, to reduce the number of input genes required. The final models were validated by leave-one-out cross-validation, in which the gene expression data of one treatment were sequentially removed from the model fitting, the parameters were re-estimated and the removed gene expression data was then used to make independent predictions of life-history variables for that treatment (Figure 1). In particular, the general cross validation criterion $(n\ D/(n - \mathrm{DoF})^2$ with n the number of data points, D the deviance and DoF the effective degress of freedom of the model) was used to estimate the smoothing parameters (method: GCV.Cp in R,[37]). All GAMs were constructed using the Gaussian distribution and the identity link function. Overall, we repeated this process three times to develop four different GAM models: one general GAM based on all available data (GEN model), and three group-specific GAMS, one based on all data from all cyanobacteria treatments (CYANO model), one based on all data from all insecticide treatments (INS model) and one based on all data from all mixture treatments (MIX model) (Figure 2).

The modules used in the GAMs were visualized using Cytoscape (version 3.4.0) and the functions and annotations of these genes were analyzed for enrichment using Fisher's tests as described in.[38]

Generally, response addition or dose addition models are used to predict the effects of mixtures.[39] However, these models cannot account for potential synergistic or antagonistic interactions. Synergistic and antagonistic interactions are determined as deviations from these models when compared observed and predicted data. Here, we develop a GAM model that accounts for interaction effects on reproduction by using the gene expression data from the individual stressors (ARTMIX model). We use the same strategy as described above and illustrated in Figure 1, applied to in silico generated gene expression profiles of the mixture treatments. These profiles are created by applying an additive model with an interaction term on the gene expression profiles of the individual stressors (e.g., $Y_{\text{in silico mix}} = Y_{\text{stressor1}} + Y_{\text{stressor2}} - Y_{\text{stressor1}} \times Y_{\text{stressor2}}$, in which Y is the log2 gene expression value as derived from the linear models described in Section 2.4). Hence, interactions can be determined without the need for comparison with observed data of actual mixture exposures. The full R code with example data is available in the Supporting Information.

3. RESULTS

We generated 144 gene expression profiles from RNA samples of *Daphnia* exposed to cyanobacteria (48 samples), insecticides (48 samples) and binary mixtures of cyanobacteria and insecticides (48 samples). These gene expression profiles were then used to construct gene networks and generalized additive models that predict effects on reproduction (Figures 1 and 2).

Based on all 144 expression profiles, we developed a gene network based on all available data that contained 4413 genes. These 4413 genes all correlated significantly with changes in reproduction following the approach in Figure 1. Of these 4413 network genes, 167 genes were ultimately included in the GEN model based on all data to predict effects on reproduction through the implementation of selection criteria as illustrated in Figure 1. Addition of more network genes did not result in any further improvement of the model following the approach described in Section 2.6 (Figure 1). Overall, the GEN model was able to explain 57.7% of the deviation Table 2). It predicted a larger effect on reproduction than observed in 38 of the 48 insecticide treatments, but predicted a smaller effect on reproduction than observed for 35 of the mixture treatments (Figure 3). An equal number of over- and underestimations by

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326

Table 2. Overview of the Different Gene Networks That Were Built Using Different Sets of Gene Expression Profiles

	stressor-specific models:		mixture models:		general model:
	CYANO	INS	MIX	ARTMIX	GEN
number of expression profiles	48	48	48	48	144
number of genes in network	1846	1189	3348	3189	4413
number of modules in GAM	3	2	3	4	4
number of genes in GAM	231	184	484	225	167
deviation explained by GAM	96.5%	87.1%	68.9%	78.9%	57.3%
R-squared of GAM	0.944	0.809	0.631	0.726	0.537
number of predictions overlapping with observations	46	46	33	40	78

the model was observed for the cyanobacteria treatments. Overall, the 95% confidence interval of predicted reproduction overlapped with observed reproduction and its standard variation for 107 out of 144 treatments (SI Figure S4, Table 2).

The group specific GAMs for insecticides and cyanobacteria, developed to predict reproduction under exposure to either insecticides or cyanobacteria, were more accurate in predicting the effects on reproduction than general GAM (Table 2, Figures 3 and 4). For both INS model and CYANO model, 95% prediction confidence intervals overlapped with the observations for 46 of the 48 treatments. For the MIX model, based on only the mixture data, the difference was smaller. Indeed, confidence intervals of predicted reproduction with the MIX model overlapped with observed reproduction for 33 of the 48 treatments versus 26 out of 48 for the GEN model (Table 2, Figure 4).

From the four developed GAM models, it is clear that reproduction under cyanobacteria and under insecticide stress can be well predicted from gene networks based on cyanobacteria or insecticide data only. The addition of gene expression profiles of mixtures or other stressors in the model development decreased the accuracy of the prediction for these individual treatments significantly. Furthermore, the gene network using all data is also 2−4 times larger (Table 2). For mixture treatments, the difference between the two models was smaller and although the MIX model based on only mixture data did result in better prediction. To improve the prediction of reproduction in mixture treatments, we developed a fifth GAM model. This ARTMIX model is based on gene networks constructed using in silico generated mixture gene expression profiles by applying an additive model with an interaction term on the gene expression profiles of the cyanobacteria treatments and the insecticide treatments as described in Section 2.6. This ARTMIX model based on in silico gene expression profiles resulted in better predictions than both the GEN model and the MIX model that both used gene expression profiles from experimental mixture data (Table 2, Figure 5). The 48 mixtures included 16 significantly antagonistic mixtures and one significantly synergistic mixture (SI Table S7). Among the eight mixtures whose predictions and observations did not have overlapping confidence intervals were four antagonistic mixtures (*Aphanizomenon* combined with chlorpyrifos and deltamethrin, *Oscillatoria* and fenoxycarb, *Cylidrospermopsis* and chlorpyrifos) and four additive mixtures. The lack of overlap between the prediction and observation for the antagonistic mixture chlorpyrifos and *Aphanizomenon* was surprising as the model was able to predict the antagonistic mixture carbaryl and *Aphanizomenon*. Given that both carbaryl and chlorpyrifos are acetylcholine esterase inhibitors, the expectation from a mechanistic point of view would be that interactions could be equally well predicted for both stressors.

Taking a closer look at the gene networks, we observed no common genes between the gene set of INS model, the CYANO model and the GEN model. A small subset of eight genes did overlap between the CYANO model and the MIX model, and 31 genes between the MIX model and the GEN model. The ARTMIX model shared five genes with the CYANO model and eight genes with the INS model but no genes with the other models. We also observed differences among network parameters between the different gene networks (SI Table S8). In particular, we observed a much lower clustering coefficient and network density for the

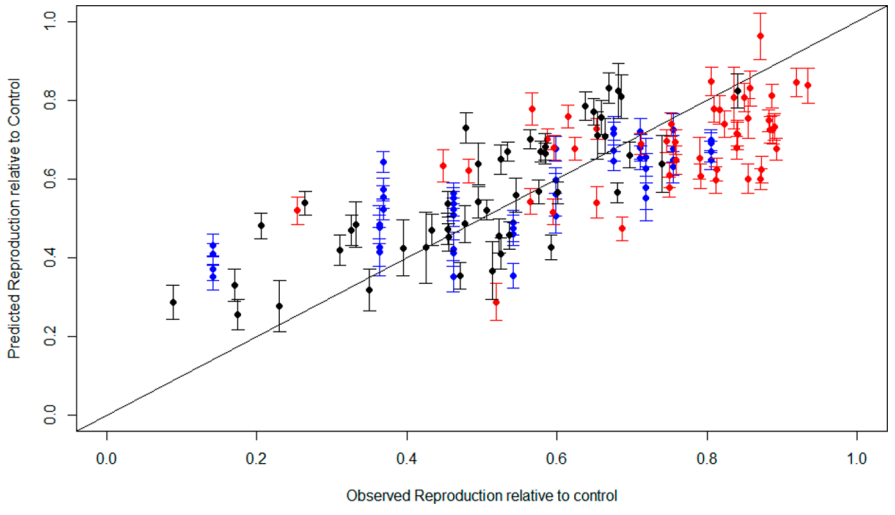

Figure 3. Observed reproduction for all treatments (cyanobacteria treatments: blue, insecticide treatments: red, binary combinations of both treatments: black) versus the predicted reproduction using a generalized additive model fitted on gene networks selected using gene expression data from all treatments (GEN model).

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326

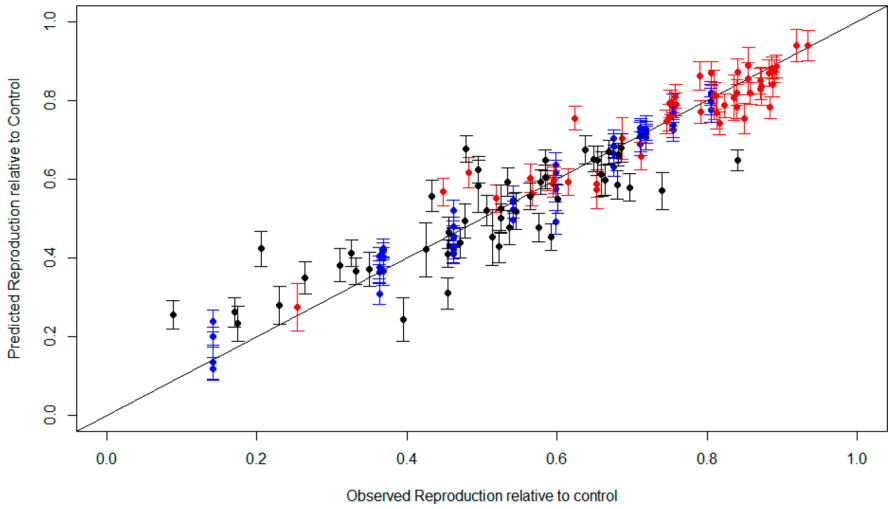

Figure 4. Observed reproduction for all treatments (cyanobacteria treatments: blue,CYANO model, insecticide treatments: red, NS model, binary combinations of both treatments: black, MIX model) versus the predicted reproduction using a generalized additive model fitted on gene networks selected using stressor specific gene expression data.

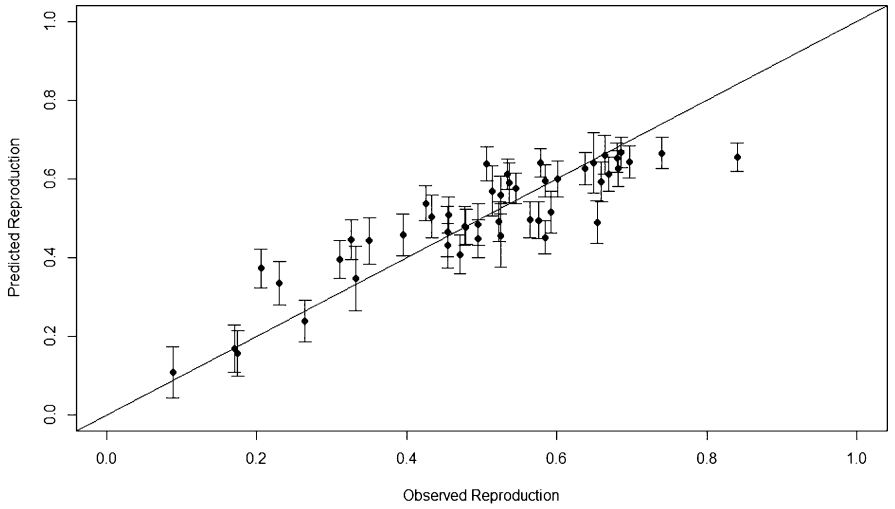

Figure 5. Observed reproduction for mixture treatments versus the predicted reproduction using a generalized additive model fitted on gene networks selected on in silico generated gene expression profiles from single stressor treatments (ARTMIX model).

network generated using all data when compared to the four other networks. This result highlights that the gene network using all data has nodes that are less connected with each other compared to the other four networks that have a higher density and a higher cluster coefficient. In contrast, the network heterogeneity was higher for the gene network using all data than for the four stressor specific networks, illustrating a more heterogeneous network containing highly connected hub genes while the majority of the nodes have few connections.

Results of simulating the performance of the five different models by random gene selection rather than gene network approach (Figure 1) highlight that the R-squared and the percentage of deviation explained by the model are much lower than the gene network approach (SI Figure S8). Even when randomly selecting genes with a significant *p*-value in the gene expression profiles, the models have lower performance than the models developed by the gene network approach (SI Figure S9).

The gene functions across the different models were very diverse. There were 571 gene functions present across all models, whereas only one gene function was present in all models: Chitinase and 438 gene functions were unique to one of the five GAM models (SI Table S9). The MIX model contained the highest number of gene functions, whereas the INS model and GEN model contained the lowest number of gene functions (SI Table S9). The average number of genes per gene function in the modelswas between 10 and 20 for all models except the GEN model. For the GEN model, this number was much higher, about 50 genes per gene function on average (SI Table S9 and Table 2). The MIX model contained the highest proportion of unique gene functions, whereas the GEN model contained the smallest proportion of unique gene functions. None of the overrepresented gene functions were present in more than one model with the exception of globins and hemoproteins which were overrepresented in both the CYANO model and the INS model (SI Table S10). Pathway analysis revealed no overrepresentation of any pathway in any of the gene networks.

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326

4. DISCUSSION

Key questions in ecology and ecotoxicology are focused on understanding how organisms respond to environmental disturbances, both from natural and anthropogenic origin. Numerous studies have described changes in life history and even gene expression patterns in response to environmental changes.[40−43] These studies are contributing to an increased mechanistic understanding of specific stress responses, but fail to quantitatively link changes in gene expression patterns to changes at life history parameters.

Here, we develop generalized additive models based on gene networks to predict effects on reproduction by using gene expression values.[36,37] We generate a gene expression data set that contained gene expression profiles of *Daphnia* exposed to cyanobacteria, insecticides, and binary combinations of the two, resulting in 144 expression profiles. Subsequent gene networks are then constructed and used as input values for generalized additive models. We observe that GAMs specifically targeted to one group of stressors are better predictors than a general GAM developed to predict effects on reproduction in all treatments. There are hardly any genes in common between the different GAMs, suggesting that the groups of stressors are very different and that the different groups of genes in the networks respond very differently to stress. The genes used in the general GAM (GEN model) are decent predictors of effects on reproduction regardless of the type of stressors. In contrast, the insecticide-specific (INS model) and cyanobacteria-specific (CYANO model) GAMs consist of genes that are good predictors of effects on reproductions specifically caused by either insecticides or cyanobacteria despite the very different modes of actions for the insecticides (SI Table S2) and the different types of cyanobacteria used (SI Table S1). This is also confirmed by the low proportion of unique gene functions in the general GAMs while the group-specific genes have a higher proportion of unique gene functions. In literature, the term general stress response genes has been used, but no clear criteria or terminology have been established to determine whether a gene can be considered a general stress response gene or not.[19,44] The current methodology with gene networks as basis for predictive modeling could provide a framework to identify general stress response genes by using defined criteria. Indeed, genes that are included in general GAMs as predictors of a wide set of environmental stressors could be considered as general stress response genes. In contrast, genes from the insecticide-specific or cyanobacteria-specific could be considered as group-specific stress response genes, responding specifically to a group of stressors. Furthermore, our results highlight these group-specific stress response genes are better quantitative predictors of reproductive effects in response to that specific group of stressors than general stress response genes.

Functional analysis of the final reduced networks, in which all genes correlated significantly with the effects on reproduction, revealed little overrepresentation of gene functions and no specific pathways were overrepresented. This result suggests that the gene networks driving changes in reproduction are most likely a construction of key genes from many different pathways and functional groups that are specifically affected by the stressors with effects on reproduction as a consequence. The lack of overlap and similarities between the insecticide specific gene network and cyanobacteria gene network, further underlines that reproduction can be perturbed through various pathways and mechanisms and that the stressor-specific GAMs are really targeted at the functional level toward a specific group of stressors. This result is in line with the adverse outcome pathway concept that has been put forward in environmental risk assessment as a mechanistic tool to evaluate the toxicity of chemicals.[45,46] Adverse outcome pathways are designed to detail the mechanistic responses from the initial first interaction between the stressor and the organism (called molecular initiating event) to the final adverse outcome.[47] Hence, they are not chemical specific, but chemicals are grouped by their molecular initiating event and key events that define the pathway.[48] Here, we put forward evidence that the adverse outcome caused by stressors that differ in these key events, for example, different insecticides, can still be well represented by a group-specific gene network that contains a specific set of genes and can be clearly linked to adverse outcomes at the organismal level. This is in line with what is referred to as AOP networks, which are groups of AOP that together represent the effect of chemical or stressor.[49] While our results provided limited mechanistic understanding on modes of action, it does provide mechanistic understanding on the gene networks contributing to adverse outcomes. Furthermore, the current GAM models allow clear quantitative links between gene expression and effects on reproduction. In contrast to individual AOPs, they do not require any mechanistic knowledge to predict organismal effect and as a consequence the developed GAMs have a more limited mechanistic value and cannot improve our understanding of the specific mechanisms of each stressor.[48]

Unsurprisingly, effects of mixtures were more difficult to predict than effects of single stressors. Yet, significant predictions could be made by applying a two way anova concept to gene expression profiles of individual stressors to generate in silico mixture profiles that can be implemented in a GAM model. These predictions are more accurate than predictions made by GAM models that directly use expression profiles of mixtures. The results highlight that effects on reproduction can be well predicted by the expression profiles of the individual stressors, including stressor combinations that result in interactions at the reproduction level. As such, these results provide a new potential method to predict the effects of mixtures, which has been recognized as one of the biggest gaps in current risk assessment strategies and regulatory approaches.[21] Currently, concentration and response addition models can predict the combined effects of stressors by using life history data of the single stressor exposures but cannot account for potential antagonistic or synergistic interactions.[21,50] Therefore, mixture effects are determined by in vivo experiments in which organisms are exposed to each mixture and the individual compounds to then determine the potential effect of the mixture.[21,50] This is a highly intensive and time-consuming process to evaluate an infinite number of potential mixture combinations whereas the results presented here could provide a significantly faster method of predictive modeling. Indeed, the current GAM models were the result of leave-one-out cross-validations in which predictions of effects on reproduction were made by using gene expression profiles of single stressors without using any life history data of the mixtures. From a mechanistic point of view, our results suggest that reproductive effects are primarily driven by genes specific for each stressor and the expression of these genes can be obscured by additional new gene expression responses in the mixture that do not significantly contribute to effects on reproduction.

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326

While these results are promising, and imply a convenient way to predict potential interactive effects of new mixtures for an infinite number of chemicals in the environment, the question remains to what extent these results can be extrapolated across different stressors, across concentration response curves, through time and across mixture complexity. Our current models focus on binary mixtures but it remains to be studied whether these models can extend to mixtures with increasing numbers of components. Indeed, environments are often contaminated with many different chemicals leading to complex mixtures which may be much more difficult to model and predict. Yet, this question is important in bridging the gap to the real world complexity. Answering these questions will require a significant amount of new molecular data for validation and evaluation. Even though an increasing number of gene expression data sets is publicly available, the lack of corresponding life history data makes it impossible at this time for us to validate our developed models in a wider context.

Nevertheless, our developed GAMs put forward a new approach to estimate and quantify effects of natural and anthropogenic stressors on aquatic organisms. By incorporating the information in gene expression profiles into gene networks and statistical models, gene expression values of specific gene sets can be quantitatively linked to higher organismal effects such as reproduction. Our results underline the potential of gene expression data as quantitative predictors of apical end points.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.7b05256.

Detailed methods on experiments, chemical and RNA analyses. Including tables of cyanobacterial strains, effect concentrations for individual insecticdes, concentrations of insecticides used in the experiments, deviation parameters for life history experiments, network parameters for the different gene networks, lists of gene functions for the different generalized additive models. Including figures on log transformed reproduction for all experiments, on predicted and observed reproduction for the different generalized additive models, figures on the performance of models generated by random sampling versus models based on gene networks (PDF)

R code for WGCNA analysis and GAM modeling, with example gene expression data and trait data (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jana.asselman@ugent.be.

ORCID

J. Asselman: 0000-0003-0185-6516

Author Contributions

J.A. and K.D.S. designed the study. J.A. and J.A.L. performed the experiments, extracted the RNA and did the microarray analysis. J.A. analyzed the results with comments and suggestions from K.D.S., M.E.P. and J.RS. J.A. has written the manuscript with comments and suggestions from K.D.S., M.E.P., and J.R.S. None of the authors have competing interests.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.A. is a postdoctoral fellow of the FWO. Funding was received from BELSPO (AquaStress project: BELSPO IAP Project P7/31), from UGent research fund (BOF15/24J/106). This research was made possible, in part, with support from the Notre Dame Genomics and Bioinformatics Core Facility through utilizing microarray facility. We thank Dieter De Coninck, Jolien De Pecker, Nancy De Saeyer, Stephen Glaholt and Leen Van Imp for the technical assistance. This research contributes to and benefits from the Daphnia Genomics Consortium.

■ REFERENCES

(1) Lawson, C. R.; Vindenes, Y.; Bailey, L.; van de Pol, M. Environmental variation and population responses to global change. *Ecol Lett.* **2015**, *18* (7), 724−736.

(2) Ehrlen, J.; Morris, W. F. Predicting changes in the distribution and abundance of species under environmental change. *Ecol Lett.* **2015**, *18* (3), 303−314.

(3) Vorosmarty, C. J.; McIntyre, P. B.; Gessner, M. O.; Dudgeon, D.; Prusevich, A.; Green, P.; Glidden, S.; Bunn, S. E.; Sullivan, C. A.; Liermann, C. R.; Davies, P. M. Global threats to human water security and river biodiversity. *Nature* **2010**, *467* (7315), 555−61.

(4) Holmstrup, M. Reprint of: The ins and outs of water dynamics in cold tolerant soil invertebrates. *J. Therm. Biol.* **2015**, *54*, 30−6.

(5) Kitano, H. Computational systems biology. *Nature* **2002**, *420* (6912), 206−10.

(6) Villeneuve, D. L.; Garcia-Reyero, N. Vision & strategy: Predictive ecotoxicology in the 21st century. *Environ. Toxicol. Chem.* **2011**, *30* (1), 1−8.

(7) Thomas, R. S.; Black, M. B.; Li, L. L.; Healy, E.; Chu, T. M.; Bao, W. J.; Andersen, M. E.; Wolfinger, R. D. Response to Incorporating Biological, Chemical, and Toxicological Knowledge Into Predictive Models of Toxicity. *Toxicol. Sci.* **2012**, *130* (2), 442−443.

(8) Thomas, R. S.; Clewell, H. J.; Allen, B. C.; Yang, L. L.; Healy, E.; Andersen, M. E. Integrating pathway-based transcriptomic data into quantitative chemical risk assessment: A five chemical case study. *Mutat. Res., Genet. Toxicol. Environ. Mutagen.* **2012**, *746* (2), 135−143.

(9) Fuller, T. F.; Ghazalpour, A.; Aten, J. E.; Drake, T. A.; Lusis, A. J.; Horvath, S. Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm. Genome* **2007**, *18* (6−7), 463−472.

(10) Yang, Y.; Han, L.; Yuan, Y.; Li, J.; Hei, N.; Liang, H. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat. Commun.* **2014**, *5*, 3231.

(11) Saris, C. G. J.; Horvath, S.; van Vught, P. W. J.; van Es, M. A.; Blauw, H. M.; Fuller, T. F.; Langfelder, P.; DeYoung, J.; Wokke, J. H. J.; Veldink, J. H.; van den Berg, L. H.; Ophoff, R. A., Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. *BMC Genomics* **2009**, *10*.40510.1186/1471-2164-10-405

(12) Zhang, S.; Ng, M. K. Gene-microRNA network module analysis for ovarian cancer. *BMC Syst. Biol.* **2016**, *10* (Suppl 4), 117.

(13) Liu, R.; Zhang, W.; Liu, Z. Q.; Zhou, H. H. Associating transcriptional modules with colon cancer survival through weighted gene co-expression network analysis. *BMC Genomics* **2017**, *18* (1), 361.

(14) Ficklin, S. P.; Luo, F.; Feltus, F. A. The association of multiple interacting genes with specific phenotypes in rice using gene coexpression networks. *Plant Physiol.* **2010**, *154* (1), 13−24.

(15) Weston, D. J.; Gunter, L. E.; Rogers, A.; Wullschleger, S. D., Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants. *BMC Syst. Biol.* **2008**, 2.1610.1186/1752-0509-2-16

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326

(16) Cassese, A.; Guindani, M.; Antczak, P.; Falciani, F.; Vannucci, M. A Bayesian Model for the Identification of Differentially Expressed Genes in Daphnia Magna Exposed to Munition Pollutants. *Biometrics* **2015**, *71* (3), 803−811.
(17) Antczak, P.; Jo, H. J.; Woo, S.; Scanlan, L.; Poynton, H.; Loguinov, A.; Chan, S.; Falciani, F.; Vulpe, C. Molecular toxicity identification evaluation (mTIE) approach predicts chemical exposure in Daphnia magna. *Environ. Sci. Technol.* **2013**, *47* (20), 11747−56.
(18) Scanlan, L. D.; Loguinov, A. V.; Teng, Q.; Antczak, P.; Dailey, K. P.; Nowinski, D. T.; Kornbluh, J.; Lin, X. X.; Lachenauer, E.; Arai, A.; Douglas, N. K.; Falciani, F.; Stapleton, H. M.; Vulpe, C. D. Gene transcription, metabolite and lipid profiling in eco-indicator daphnia magna indicate diverse mechanisms of toxicity by legacy and emerging flame-retardants. *Environ. Sci. Technol.* **2015**, *49* (12), 7400−10.
(19) Roelofs, D.; Aarts, M. G. M.; Schat, H.; van Straalen, N. M. Functional ecological genomics to demonstrate general and specific responses to abiotic stress. *Funct Ecol* **2008**, *22* (1), 8−18.
(20) Heugens, E. H.; Hendriks, A. J.; Dekker, T.; van Straalen, N. M.; Admiraal, W. A review of the effects of multiple stressors on aquatic organisms and analysis of uncertainty factors for use in risk assessment. *Crit. Rev. Toxicol.* **2001**, *31* (3), 247−84.
(21) Altenburger, R.; Backhaus, T.; Boedeker, W.; Faust, M.; Scholze, M. Simplifying complexity: Mixture toxicity assessment in the last 20 years. *Environ. Toxicol. Chem.* **2013**, *32* (8), 1685−7.
(22) Shaw, J. R.; Colbourne, J. K.; Davey, J. C.; Glaholt, S. P.; Hampton, T. H.; Chen, C. Y.; Folt, C. L.; Hamilton, J. W. Gene response profiles for Daphnia pulex exposed to the environmental stressor cadmium reveals novel crustacean metallothioneins. *BMC Genomics* **2007**, *8*, 477.
(23) Asselman, J.; Hochmuth, J. D.; De Schamphelaere, K. A. C. A comparison of the sensitivities of Daphnia magna and Daphnia pulex to six different cyanobacteria. *Harmful Algae* **2014**, *39*, 1−7.
(24) Asselman, J.; Meys, J.; Waegeman, W.; De Baets, B.; De Schamphelaere, K. A. C. Combined Exposure to Cyanobacteria and Carbaryl Results in Antagonistic Effects on the Reproduction of Daphnia Pulex. *Environ. Toxicol. Chem.* **2013**, *32* (9), 2153−2158.
(25) Asselman, J.; Janssen, C. R.; Smagghe, G.; De Schamphelaere, K. A. C. Ecotoxicity of binary mixtures of Microcystis aeruginosa and insecticides to Daphnia pulex. *Environ. Pollut.* **2014**, *188*, 56−63.
(26) Asselman, J.; Pfrender, M. E.; Lopez, J. A.; De Coninck, D. I. M.; Janssen, C. R.; Shaw, J. R.; De Schamphelaere, K. A. C. Conserved transcriptional responses to cyanobacterial stressors are mediated by alternate regulation of paralogous genes in Daphnia. *Mol. Ecol.* **2015**, *24* (8), 1844−1855.
(27) De Coninck, D. I. M.; De Schamphelaere, K. A. C.; Jansen, M.; De Meester, L.; Janssen, C. R. Interactive effects of a bacterial parasite and the insecticide carbaryl to life-history and physiology of two Daphnia magna clories differing in carbaryl sensitivity. *Aquat. Toxicol.* **2013**, *130*, 149−159.
(28) Royston, J. P. An Extension of Shapiro and Wilk-W Test for Normality to Large Samples. *Appl. Stat-J. Roy St C* **1982**, *31* (2), 115−124.
(29) Fox, J. *Applied Regression Analysis and Generalized Linear Models*, 2nd ed.; Sage Publications, 2008.
(30) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J. Roy Stat Soc. B Met* **1995**, *57* (1), 289−300.
(31) Faust, M.; Altenburger, R.; Backhaus, T.; Blanck, H.; Boedeker, W.; Gramatica, P.; Hamer, V.; Scholze, M.; Vighi, M.; Grimme, L. H. Joint algal toxicity of 16 dissimilarly acting chemicals is predictable by the concept of independent action. *Aquat. Toxicol.* **2003**, *63* (1), 43−63.
(32) Colbourne, J. K.; Pfrender, M. E.; Gilbert, D.; Thomas, W. K.; Tucker, A.; Oakley, T. H.; Tokishita, S.; Aerts, A.; Arnold, G. J.; Basu, M. K.; Bauer, D. J.; Caceres, C. E.; Carmel, L.; Casola, C.; Choi, J. H.; Detter, J. C.; Dong, Q. F.; Dusheyko, S.; Eads, B. D.; Frohlich, T.; Geiler-Samerotte, K. A.; Gerlach, D.; Hatcher, P.; Jogdeo, S.; Krijgsveld, J.; Kriventseva, E. V.; Kultz, D.; Laforsch, C.; Lindquist, E.; Lopez, J.; Manak, J. R.; Muller, J.; Pangilinan, J.; Patwardhan, R. P.; Pitluck, S.; Pritham, E. J.; Rechtsteiner, A.; Rho, M.; Rogozin, I. B.; Sakarya, O.; Salamov, A.; Schaack, S.; Shapiro, H.; Shiga, Y.; Skalitzky, C.; Smith, Z.; Souvorov, A.; Sung, W.; Tang, Z. J.; Tsuchiya, D.; Tu, H.; Vos, H.; Wang, M.; Wolf, Y. I.; Yamagata, H.; Yamada, T.; Ye, Y. Z.; Shaw, J. R.; Andrews, J.; Crease, T. J.; Tang, H. X.; Lucas, S. M.; Robertson, H. M.; Bork, P.; Koonin, E. V.; Zdobnov, E. M.; Grigoriev, I. V.; Lynch, M.; Boore, J. L. The Ecoresponsive Genome of Daphnia pulex. *Science* **2011**, *331* (6017), 555−561.
(33) Gentleman, R. C.; Carey, V. J.; Bates, D. M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; Hornik, K.; Hothorn, T.; Huber, W.; Iacus, S.; Irizarry, R.; Leisch, F.; Li, C.; Maechler, M.; Rossini, A. J.; Sawitzki, G.; Smith, C.; Smyth, G.; Tierney, L.; Yang, J. Y.; Zhang, J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **2004**, *5* (10), R80.
(34) Smyth, G. K., Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*, Article3.110.2202/1544-6115.1027
(35) Ye, Z.; Xu, S.; Spitze, K.; Asselman, J.; Jiang, X.; Ackerman, M. S.; Lopez, J.; Harker, B.; Raborn, R. T.; Thomas, W. K.; Ramsdell, J.; Pfrender, M. E.; Lynch, M. A New Reference Genome Assembly for the Microcrustacean Daphnia pulex. *G3: Genes, Genomes, Genet.* **2017**, *7* (5), 1405−1416.
(36) Langfelder, P.; Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* **2008**, *9*, 559.
(37) Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.* **2004**, *99* (467), 673−686.
(38) Asselman, J.; De Coninck, D. I. M.; Glaholt, S.; Colbourne, J. K.; Janssen, C. R.; Shaw, J. R.; De Schamphelaere, K. A. C. Identification of Pathways, Gene Networks, and Paralogous Gene Families in Daphnia pulex Responding to Exposure to the Toxic Cyanobacterium Microcystis aeruginosa. *Environ. Sci. Technol.* **2012**, *46* (15), 8448−8457.
(39) Jonker, M. J.; Svendsen, C.; Bedaux, J. J. M.; Bongers, M.; Kammenga, J. E. Significance testing of synergistic/antagonistic, dose level-dependent, or dose ratio-dependent effects in mixture dose-response analysis. *Environ. Toxicol. Chem.* **2005**, *24* (10), 2701−2713.
(40) Latta, L. C.; Weider, L. J.; Colbourne, J. K.; Pfrender, M. E. The evolution of salinity tolerance in Daphnia: a functional genomics approach. *Ecol Lett.* **2012**, *15* (8), 794−802.
(41) Dubansky, B.; Whitehead, A.; Miller, J. T.; Rice, C. D.; Galvez, F. Multitissue Molecular, Genomic, and Developmental Effects of the Deepwater Horizon Oil Spill on Resident Gulf Killifish (Fundulus grandis). *Environ. Sci. Technol.* **2013**, *47* (10), 5074−5082.
(42) Poynton, H. C.; Lazorchak, J. M.; Impellitteri, C. A.; Blalock, B. J.; Rogers, K.; Allen, H. J.; Loguinov, A.; Heckrnan, J. L.; Govindasmawy, S. Toxicogenomic Responses of Nanotoxicity in Daphnia magna Exposed to Silver Nitrate and Coated Silver Nanoparticles. *Environ. Sci. Technol.* **2012**, *46* (11), 6288−6296.
(43) Heckmann, L. H.; Sibly, R. M.; Connon, R.; Hooper, H. L.; Hutchinson, T. H.; Maund, S. J.; Hill, C. J.; Bouetard, A.; Callaghan, A. Systems biology meets stress ecology: linking molecular and organismal stress responses in Daphnia magna. *Genome Biology* **2008**, *9*, R40.
(44) Even, N.; Devaud, J. M.; Barron, A. B. General Stress Responses in the Honey Bee. *Insects* **2012**, *3* (4), 1271−98.
(45) Ankley, G. T.; Bennett, R. S.; Erickson, R. J.; Hoff, D. J.; Hornung, M. W.; Johnson, R. D.; Mount, D. R.; Nichols, J. W.; Russom, C. L.; Schmieder, P. K.; Serrrano, J. A.; Tietge, J. E.; Villeneuve, D. L. Adverse Outcome Pathways: A Conceptual Framework to Support Ecotoxicology Research and Risk Assessment. *Environ. Toxicol. Chem.* **2010**, *29* (3), 730−741.
(46) Fay, K. A.; Villeneuve, D. L.; LaLone, C. A.; Song, Y.; Tollefsen, K. E.; Ankley, G. T. Practical approaches to adverse outcome pathway development and weight-of-evidence evaluation as illustrated by ecotoxicological case studies Environmental Toxicology and Chemistry Volume 36, Issue 6. *Environ. Toxicol. Chem.* **2017**, *36*, 1429−1449.

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326

(47) Villeneuve, D. L.; Crump, D.; Garcia-Reyero, N.; Hecker, M.; Hutchinson, T. H.; LaLone, C. A.; Landesmann, B.; Lettieri, T.; Munn, S.; Nepelska, M.; Ottinger, M. A.; Vergauwen, L.; Whelan, M. Adverse Outcome Pathway (AOP) Development I: Strategies and Principles. *Toxicol. Sci.* **2014**, *142* (2), 312−320.
(48) Perkins, E. J.; Chipman, J. K.; Edwards, S.; Habib, T.; Falciani, F.; Taylor, R.; Van Aggelen, G.; Vulpe, C.; Antczak, P.; Loguinov, A. Reverse Engineering Adverse Outcome Pathways. *Environ. Toxicol. Chem.* **2011**, *30* (1), 22−38.
(49) Knapen, D.; Vergauwen, L.; Villeneuve, D. L.; Ankley, G. T. The potential of AOP networks for reproductive and developmental toxicity assay development. *Reprod. Toxicol.* **2015**, *56*, 52−55.
(50) Altenburger, R.; Boedeker, W.; Faust, M.; Grimme, L. H. Regulations for combined effects of pollutants: Consequences from risk assessment in aquatic toxicology. *Food Chem. Toxicol.* **1996**, *34* (11−12), 1155−1157.

DOI: 10.1021/acs.est.7b05256
Environ. Sci. Technol. 2018, 52, 317−326