UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

The Toxicogenome of Hyalella azteca

Poynton, Helen C.; Colbourne, John K.; Hasenbein, Simone; Benoit, Joshua B.; Sepulveda, Maria S.; Poelchau, Monica F.; Hughes, Daniel S.T.; Murali, Shwetha C.; Chen, Shuai; Glastad, Karl M.; Goodisman, Michael A.D.; Werren, John H.; Vineis, Joseph H.; Bowen, Jennifer L.; Friedrich, Markus; Jones, Jeffery; Robertson, Hugh M.; Feyereisen, René; Mechler-Hickson, Alexandra; Mathers, Nicholas

DOI:
10.1021/acs.est.8b00837

License:
None: All rights reserved

Document Version
Peer reviewed version

Citation for published version (Harvard):
Poynton, HC, Colbourne, JK, Hasenbein, S, Benoit, JB, Sepulveda, MS, Poelchau, MF, Hughes, DST, Murali, SC, Chen, S, Glastad, KM, Goodisman, MAD, Werren, JH, Vineis, JH, Bowen, JL, Friedrich, M, Jones, J, Robertson, HM, Feyereisen, R, Mechler-Hickson, A, Mathers, N, Lee, CE, Biales, A, Johnston, JS, Wellborn, GA, Rosendale, AJ, Cridge, AG, Munoz-Torres, MC, Bain, PA, Manny, AR, Major, KM, Lambert, FN, Vulpe, CD, Tuck, P, Blalock, BJ, Lin, YY, Smith, ME, Ochoa-Acuña, H, Chen, MJM, Childers, CP, Qu, J, Dugan, S, Lee, SL, Chao, H, Dinh, H, Han, Y, Doddapaneni, H, Worley, KC, Muzny, DM, Gibbs, RA & Richards, S 2018, 'The Toxicogenome of Hyalella azteca: A Model for Sediment Ecotoxicology and Evolutionary Toxicology', *Environmental Science and Technology*, vol. 52, no. 10, pp. 6009-6022. https://doi.org/10.1021/acs.est.8b00837

Link to publication on Research at Birmingham portal

Publisher Rights Statement:
This document is the Accepted Manuscript version of a Published Work that appeared in final form in Environmental Science and Technology, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see DOI: 10.1021/acs.est.8b00837. For non-commercial use only.

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 26. Apr. 2024

The Toxicogenome of *Hyalella azteca*: a model for sediment ecotoxicology and evolutionary toxicology

Helen C. Poynton[1,*], Simone Hasenbein[2], Joshua B. Benoit[3], Maria S. Sepulveda[4], Monica F. Poelchau[5], Daniel S.T. Hughes[6], Swetha C. Murali[6], Shuai Chen[4,7], Karl M. Glastad[8,9], Michael A. D. Goodisman[9], John H. Werren[10], Joseph H. Vineis[11], Jennifer L. Bowen[11], Markus Friedrich[12], Jeffery Jones[12], Hugh M. Robertson[13], René Feyereisen[14], Alexandra Mechler-Hickson[15], Nicholas Mathers[15], Carol Eunmi Lee[15], John K. Colbourne[16], Adam Biales[17], J. Spencer Johnston[18], Gary A. Wellborn[19], Andrew J. Rosendale[3], Andrew G. Cridge[20], Monica C. Munoz-Torres[21], Peter A. Bain[22], Austin R. Manny[23], Kaley M. Major[1], Faith Lambert[24], Chris D. Vulpe[24], Padrig Tuck[1], Bonnie Blalock[1],Yu-Yu Lin[25], , Mark E. Smith[26], Hugo Ochoa-Acuña[4], Mei-ju May Chen[25], Christopher P. Childers[5], Jiaxin Qu[6], Shannon Dugan[6], Sandra L. Lee[6], Hsu Chao[6], Huyen Dinh[6], Yi Han[6], HarshaVardhan Doddapaneni[6], Kim C. Worley[6, 27], Donna M. Muzny[6], Richard A. Gibbs[6], Stephen Richards[6]

Affiliations

1. School for the Environment, University of Massachusetts Boston, Boston, MA 02125 United States
2. Aquatic Systems Biology Unit, Technical University of Munich, D-85354 Freising, Germany
3. Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221 United States
4. Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907 United States
5. Agricultural Research Service, National Agricultural Library, USDA, Beltsville, MD 20705 United States
6. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030 United States
7. OmicSoft Corporation, Cary, NC 27513 United States
8. Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 United States
9. School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332 United States
10. Biology Department, University of Rochester, Rochester, NY 14627 United States
11. Department of Marine and Environmental Sciences, Marine Science Center, Northeastern University, Nahant, MA 01908 United States
12. Wayne State University, Detroit MI 48202 United States
13. Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801 United States
14. Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg, Denmark
15. Center of Rapid Evolution (CORE) and Department of Integrative Biology, University of Wisconsin, Madison WI 53706 USA
16. School of Biosciences, University of Birmingham, Birmingham, B15 2TT UK
17. National Exposure Research Laboratory, United States Environmental Protection Agency, Cincinnati, OH 45268 United States
18. Department of Entomology, Texas A&M University, College Station, TX 77843 USA
19. Department of Biology, University of Oklahoma, Norman, OK 73019 United States

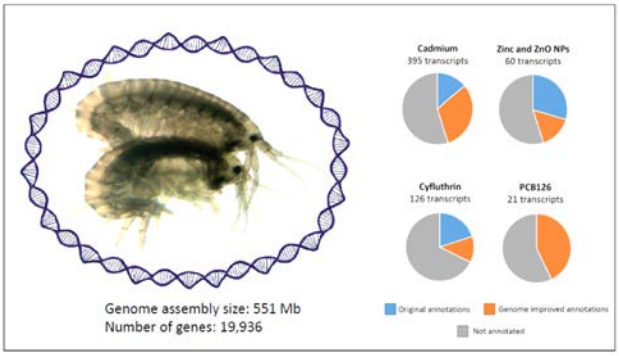
20. Laboratory for Evolution and Development, Department of Biochemistry, University of Otago, Dunedin, 9054 New Zealand
21. Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 United States
22. Commonwealth Scientific and Industrial Research Organisation (CSIRO), Urrbrae SA, 5064 Australia
23. Department of Microbiology & Cell Science, University of Florida, Gainesville, FL 32611 United States
24. Center for Environmental and Human Toxicology, University of Florida, Gainesville, FL 32611 United States
25. Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, 10617 Taiwan
26. McConnell Group, Cincinnati, OH 45268, United States
27. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030 United States

*Corresponding author: Helen Poynton
School for the Environment, University of Massachusetts Boston
100 Morrissey Blvd., Boston, MA 02125, USA
helen.poynton@umb.edu

Abstract

Hyalella azteca is a cryptic species complex of epibenthic amphipods of interest to ecotoxicology and evolutionary biology. It is the primary crustacean used in North America for sediment toxicity testing and an emerging model for molecular ecotoxicology. To provide molecular resources for sediment quality assessments and evolutionary studies, we sequenced, assembled, and annotated the genome of the *H. azteca* US Lab Strain. The genome quality and completeness is comparable with other ecotoxicological model species. Through targeted investigation and use of gene expression data sets of *H. azteca* exposed to pesticides, metals, and other emerging contaminants, we annotated and characterized the major gene families involved in sequestration, detoxification, oxidative stress, and toxicant response. Our results revealed gene loss related to light sensing, but a large expansion in chemoreceptors, likely underlying sensory shifts necessary in their low light habitats. Gene family expansions were also noted for cytochrome P450 genes, cuticle proteins, ion transporters, and include recent gene duplications in the metal sequestration protein, metallothionein. Mapping of differentially-expressed transcripts to the genome significantly increased the ability to functionally annotate toxicant responsive genes. The *H. azteca* genome will greatly facilitate development of genomic tools for environmental assessments and promote an understanding of how evolution shapes toxicological pathways with implications for environmental and human health.

TOC art:


Cadmium
395 transcripts
Zinc and ZnO NPs
60 transcripts
Cyfluthrin
126 transcripts
PCB126
21 transcripts
Original annotations
Genome improved annotations
Not annotated
Genome assembly size: 551 Mb
Number of genes: 19,936

Introduction

Sediment quality assessments serve as a metric for overall habitat integrity of freshwater ecosystems. Sediments provide a foundation for aquatic food webs, providing habitat for invertebrate species including crustaceans and insect larvae. However, they also concentrate pollution over time, especially hydrophobic contaminants that sorb to sediments, leading to bioaccumulation in the food web.[1] While concentrations of a few legacy contaminants are declining in the United States due to regulatory efforts;[2] chemicals designed as their replacements are becoming emerging contaminants and their levels are increasing.[3] This is particularly true of newer generation pesticides, which have become problematic in urban areas,[4] and complex mixtures of pharmaceuticals and personal healthcare products.[5]

Hyalella azteca is a freshwater crustacean (Malacostraca: Amphipoda) that lives near the sediment surface, burrowing in sediment and scavenging on leaf litter, algae, and detritus material on the sediment surface.[6] The amphipod's nearly continuous contact with sediment, rapid generation time, and high tolerance to changes in temperature and salinity has made *H. azteca* an ideal species for assessing toxicity[7, 8] and the bioavailability of sediment contaminants.[9] Given its ecology and expansive distribution, *H. azteca* provides an important window into sediment toxicant exposure and is a foundational trophic link to vertebrates as prey.[10, 11]

The *H. azteca* species complex represents one of the most abundant and broadly distributed amphipods in North America. It was originally characterized as a single cosmopolitan species, but life history and morphological differences of *H. azteca* from different locations suggest that they comprise a species complex.[12-15] Indeed, phylogeographic analyses have resolved several different species,[16-20] which have diverged in North America over the past 11 million years.[19] **Figure 1** shows the distribution of seven of the best characterized species within the complex. Even as the species have diverged, convergent evolution appears to be occurring due to similarity between geographically dispersed habitats providing an interesting study system for evolutionary biology.[21] For example, several populations representing multiple species groups have independently evolved genetic resistance to pyrethroid insecticides through mutations to the voltage-gate sodium channel (VGSC).[22]

Within ecotoxicology, new strategies are being promoted to address the magnitude and wide range of effects elicited by chemicals and deficiencies in current toxicity testing approaches (e.g., National Research Council[23]). These strategies include developing adverse outcome pathway models that connect 'key events' that are predictive of harmful results, from

molecular perturbations to ecologically relevant effects.[24, 25] In addition, comparative toxicogenomic approaches to identify evolutionarily conserved toxicological pathways[26, 27] and target sites[28] enable cross-species predictions of adverse effects.

To ensure that sediment toxicity testing remains current with these emerging approaches in environmental health assessments, we set out to investigate the toxicogenome of *H. azteca* (US Lab Strain), or the complete set of genes involved in toxicological pathways and stress response. As an often-used model, and the most widely used species within the complex for ecotoxicology, a relatively large amount of toxicity data has been amassed for the US Lab Strain of *H. azteca*. Detailed characterization of the toxicogenome creates opportunities to reinterpret and exploit existing toxicity data to better predict risk posed to the environment by chemicals. Here we describe the genome of *H. azteca* with particular emphasis on genes related to key toxicological targets and pathways including detoxification, stress response, developmental and sensory processes, and ion transport. In addition, we begin the process of creating a 'gene ontology' for environmental toxicology using *H. azteca* by annotating the function of genes responsive to model sediment contaminants, thereby shinning light on the toxicogenome underlying adverse outcome pathways.

Methods:

***Hyalella azteca* strain, inbreeding, and genomic DNA (gDNA) extraction**

Hylalella azteca (US Laboratory Strain[29]) cultures were reared according to standard test conditions.[10] These organisms have been maintained by the US EPA since their original collection by A. Nebeker (ca. 1982) from a stream near Corvallis, Oregon.[16] They share highest genetic similarity with populations collected in Florida and Oklahoma (**Figure 1**, Clade 8)[29, 30] and recently in California.[22] Several lines of full sibling matings were maintained for four generations, after which all lines were unable to produce offspring, likely due to inbreeding depression.[31] Twenty animals including both males and females from a single inbred line were collected and gDNA extracted from individual *H. azteca* using the Qiagen DNeasy® Blood & Tissue Kit (Qiagen, Germantown, MD) with slight modifications.[22] Because of the gDNA quantities required, multiple individuals were pooled for library construction.

gDNA sequencing, assembly, and annotation

H. azteca is one of 30 arthropod species sequenced as part of a pilot project for the i5K Arthropod Genomes Project at the Baylor College of Medicine Human Genome Sequencing

Center, Houston, Texas, USA. Sequencing was performed on Illumina HiSeq2000s (Casava v. 1.8.3_V3) generating 100 bp paired end reads. The amount of sequences generated from each of four libraries (nominal insert sizes 180 bp, 500 bp, 3 kb and 8 kb) is noted in **Table S1** with NCBI SRA accessions. See supporting information (**S1**) for more details on library preparation. Reads were assembled using ALLPATHS-LG (v35218)[32] and further scaffolded and gap-filled using in-house tools Atlas-Link (v.1.0) and Atlas gap-fill (v.2.2) (https://www.hgsc.bcm.edu/software/). This yielded an initial assembly (HAZT_1.0; **Table S1;** NCBI Accession GCA_000764305.1) of 1.18 Gb (596.68 Mb without gaps within scaffolds), compared with genome size of 1.05 Gb determined by flow cytometry (see **SI** for methods). To improve assembly contiguity, we used the Redundans[33] assembly tool. With Redundans using standard parameters, HAZT_1.0 scaffolds and all Illumina input reads given to ALLPATHS-LG when producing HAZT_1.0 as data inputs, generated a new assembly (HAZT_2.0, **Table S1;** NCBI accession GCA_000764305.2) of 550.9 Mb (548.3 Mb without gaps within scaffolds).

The HAZT_1.0 genome assembly was subjected to automatic gene annotation using a Maker 2.0 annotation pipeline tuned specifically for arthropods. The core of the pipeline was a Maker 2 instance,[34] modified slightly to enable efficient running on our computational resources. See supporting information for additional details. The automated gene sets are available from the BCM-HGSC website (https://hgsc.bcm.edu/arthropods/hyalella-azteca-genome-project), Ag Data Commons[35] and the National Agricultural Library (https://i5k.nal.usda.gov/Hyalella_azteca) where a web-browser of the genome, annotations, and supporting data is accessible. The Hazt_2.0 assembly was annotated by the automated NCBI Eukaryotic Genome Annotation Pipeline[36] and is available from NCBI (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Hyalella_azteca/100/).

Manual annotation and official gene set generation

Automated gene prediction greatly facilitates the generation of useful genomic annotations for down-stream research; however, producing accurate, high-quality genome annotations remains a challenge.[37] Manual correction of gene models generated through automated analyses – also known as manual annotation can provide improved resources for downstream projects.[38] The *Hyalella* genome consortium recruited 25 annotators to improve gene models predicted from the genome assembly scaffolds,[39] adhering to a set of rules and guidelines during the manual annotation process (https://i5k.nal.usda.gov/content/rules-web-apollo-annotation-i5k-pilot-project). Manual annotation occurred via the Apollo software[38], which allows users to annotate collaboratively on the web via the JBrowse genome browser, version 1.0.4[40]

(https://apollo.nal.usda.gov/hyaazt/jbrowse/). Two transcriptomes (see below) were provided as external evidence for manual annotation and as an additional resource to search for missing genes. After manual annotation, models were exported from Apollo and screened for general formatting and curation errors (see https://github.com/NAL-i5K/I5KNAL_OGS/wiki). Models that overlapped with potential bacterial contaminants were removed. Bacterial contamination included regions identified via the procedure outlined in the supporting information (**S1, S2**) as well as potential contamination identified by NCBI (Terence Murphy, personal communication). The remaining corrected models were then merged with MAKER gene predictions HAZTv.0.5.3 and miRNA predictions (see supporting information, **S3**.) into a non-redundant gene set, OGSv1.0;[41] for details on the merge procedure see https://github.com/NAL-i5K/I5KNAL_OGS/wiki/Merge-phase). The manual annotation process generated 911 corrected gene models, including 875 mRNAs and 46 pseudogenes. All annotations are available for download at the i5k Workspace@NAL (https://i5k.nal.usda.gov/Hyalella_azteca).[42] Additional details pertaining to the annotation of specific gene families can be found with the annotation reports of the supporting information (**S4**).

RNA sequencing and transcriptome libraries

Two sets of transcriptomic data were generated from non-exposed *H. azteca* to assist in gene prediction, and were recently published as part of a *de novo* transcriptome assembly project to identify peptide hormones[43] (see **S1** for details). RNAseq reads from this transcriptome project were aligned to the *H. azteca* genome scaffolds (HAZT_1.0) using TopHat 2.0.14 with bowtie 2-2.1.0 and SAMtools 1.2. Overall mapping rate was 70.2%. Resulting BAM files were transferred to NAL and added to the Apollo genome browser.

Gene expression data sets

To assist in the manual annotation of genes related to toxicant stress, additional gene expression data sets were utilized, consisting of differentially expressed genes identified by microarray analysis following exposure to model pollutants (see **Table 1**). Two of these data sets were previously published[30, 44] while a third data set is new. Details related to these microarray experiments can be found in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo; Accession number: GPL17458) and the supporting information (**S1**). Contigs corresponding to the differentially expressed microarray probes were aligned to the genome using Blastn within the Apollo genome browser. Genes were manually annotated

in the areas of the genome where the contigs aligned using available MAKER and AUGUSTUS gene models and RNAseq reads to correct exon and intron boundaries if needed.

Table 1: Model toxicants and exposure concentrations for microarray gene expression analysis

Pollutant class	Chemical	Endpoints	Concentrations	References
Heavy Metals	Cadmium	LC_{10}	5.5 μg/L	*this study*
	Zinc	$^1/_{10}$ LC_{50}	25 μg/L	Poynton et al. (2013)[44]
		LC_{25}	104 μg/L	
Insecticide	Cyfluthrin	NOEC	1 ng/L	Weston et al. (2013)[30]
Nanomaterial	ZnO NP	$^1/_{10}$ LC_{50}	18 μg/L	Poynton et al. (2013)[44]
		LC_{25}	65 μg/L	
Organic Pollutant	PCB126	----	7.0 μg/L	*this study*

Results and Discussion:

Description of the *Hyalella azteca* Genome

Analysis of the *H. azteca* US Lab strain genome size using flow cytometry gave an average genome size of 1.05 Gb (females 1C = 1045 +/- 8.7 Mb; males 1C = 1061 +/- 10.2 Mb). Our genome size is significantly smaller than recent estimates from *H. azteca* representing different species groups collected throughout North America, which have been shown to vary by a factor of two.[45] Following the strategy of the i5k pilot project, we generated an assembly of 1.18 Gb (Hazt_1.0; **Table S1**). However, because of the high gap fraction and an assembly size larger than the experimentally determined genome size, a second assembly of 550.9 Mb was later generated from same sequencing data using Redundans (Hazt_2.0).[33] This assembly greatly improved contig N50 and produced a more complete gene set (**Figure 2**, **Table S1**). Because of the timing of the availability of Hazt_2.0, most of the manual annotations were performed using Hazt_1.0; therefore, both assemblies are presented here. RNAseq reads[43] mapped equally well to both genome assemblies (**Figure 2E**), but significantly more reads mapped to the improved gene models of the Hazt_2.0 assembly (**Figure 2F**). BUSCO analysis[46] was performed on both *H. azteca* genome assemblies (**Figure 2C**) and predicted gene sets (**Figure 2D**) to assess the completeness of the genome. Hazt_2.0 contains a higher percentage (Genome = 91.0%, Gene set = 94.2%) of complete BUSCOs in contrast to the

Hazt_1.0 assembly (Genome 85.3%, Gene set =67.6%). When compared to other genomes of ecotoxicological relevance,[47-53] the Hazt_2.0 showed comparable quality, ranking fourth in completeness out of the eight genomes assessed.

Associated Bacteria and Lateral Gene Transfers –

Using two complementary approaches to screen the *H. azteca* genome for bacterial contaminants,[54, 55] we recovered two draft bacterial genomes, or metagenome assembled genomes (MAGs), and evidence of lateral gene transfers (LGTs) (**S2**). MAG1 is a flavobacterium with distant affinities to currently sequenced bacteria (92% 16S rRNA identity to the most closely identified genera Chishuiella and Empedobacter). MAG2 is a related to bacteria in the genus *Ideonella* (98% 16S rRNA identity to *I. paludis*), of which some are able to degrade plastics,[56] in particular an isolate from the wax moth *Galleria mellonella*.[57] Whether these bacteria are close associates of *H. azteca* or components of their diet is not currently known, but their interesting gene repertoires could be relevant to *H. azteca* ecotoxicology. An analysis of broad functional categories indicates that both genomes contain multiple genes related to metal detoxification and resistance to toxins (i.e., antibiotics) and possibly other organic pollutants (**Figure S2.2**). In addition, strong LGT candidates from Rickettsia-like bacteria were found on five genome scaffolds (**Table S2.2**).

Genome methylation and microRNAs –

We performed two genome-wide analyses to characterize DNA methylation patterns in the genome and characterize the full complement of *H. azteca* microRNAs. DNA methylation is an epigenetic mechanism by which a methyl group (CH_3) binds to DNA that may alter gene expression.[58] Because methylated cytosines tend to mutate to thymines over evolutionary time, we used signatures of CpG dinucleotide depletion in the *H. azteca* genome to uncover putative patterns of DNA methylation. Our analyses showed that *H. azteca* possesses strong indications of genomic DNA methylation, including CpG depletion of a subset of genes (**Figure S3.1 A-B**) and presence of the key DNA methyltransferase enzymes, DNMT1 and DNMT3 (**Table S3.1**). Furthermore, genes with lower levels of CpG observed/expected (CpG o/e; i.e. putatively methylated) displayed a strong positional bias in CpG depletion (**Figure S3.1 C**, black line), with 5' regions of these genes being considerably depleted of CpGs. In contrast, genes with higher CpG o/e displayed no such positional bias (**Figure S3.1 C**, grey line). Several insects where DNA methylation has been empirically profiled at the single-base level possess such patterns[59]

suggesting *H. azteca* has similar patterns of DNA methylation as most insects (but see: Glastad et al.[60]).

MicroRNAs (miRNAs) are a family of short non-coding RNAs (~22 nt in length) that play critical roles in post-translational gene regulation.[61] Recent research revealed that miRNAs are involved in aquatic crustaceans' response to environmental stressors (e.g. hypoxia and cadmium exposure),[62, 63] which makes miRNAs promising biomarkers for future aquatic toxicological research. We predicted *H. azteca* miRNAs based on sequence homology and hairpin structure identification. A total of 1,261 candidate miRNA coding sites were identified by BLAST. After hairpin structure identification, we predicted 148 *H. azteca* miRNAs, which include several highly conserved miRNAs (e.g. miR-9 and let-7 family) (**Table S3.2**, sequences available in **S6.1**). Several Cd-responsive miRNAs in *D. pulex* (miR-210, miR-71 and miR-252)[62] were also predicted in *H. azteca*, suggesting a conserved role of these miRNAs. This number of predicted miRNAs is comparable to what has been reported for other arthropods (**Figure S3.2**).

Functional Annotation

The manual annotation of the *H. azteca* US Lab strain genome resulted in the characterization of 13 different gene families (see detailed annotation reports in supporting information **S4.1-S4.13**). Given its importance in ecological and ecotoxicological studies, a particular focus was given to genes involved in environmental sensing (chemoreceptors and opsins), detoxification and response to stress (cytochrome P450s, cuticle proteins, glutathione peroxidases, glutathione S-transferases, heat shock proteins, and metallothionein proteins), as well as genes involved in important toxicological pathways (ion transporters, early development genes, insecticide target genes, and nuclear receptors). Here we highlight significant findings of gene expansions or contractions as well as the characterization of genes of particular toxicological importance.

Environmental Sensing –

To better understand how *H. azteca* interacts with its environment, we annotated genes involved in light and chemical sensing. Arthropods deploy a diversity of light sensing mechanisms including light capture in photoreceptor cells through the expression of opsins: light-sensitive, seven-transmembrane G-protein coupled opsin receptor proteins. Our survey of the *H. azteca* genome revealed only three opsin genes, a middle wavelength-sensitive subfamily 1 (MWS1) opsin and two belonging to the long wavelength-sensitive subfamily (LWS)

opsins (**Figure S4.13.1**, **Table S4.13.1**). Maximum likelihood analysis with a subset of closely related malacostracan LWS opsins moderately supports the two *H. azteca* LWS opsins as 1:1 orthologs of the two LWS opsins previously reported for *Gammarus minus*.[64] Thus, the LWS opsin duplicate pair conserved in *H. azteca* and *G. minus* is likely ancient, predating at least the origin of amphipod Crustacea. The *H. azteca* MWS opsin by contrast, represents the first reported amphipod MWS opsin and is distinct from currently known malacostracan MWS opsins. These findings suggest that amphipod crustaceans are equipped with a minimally diversified set of three opsin genes and implies gene family losses for all of the non-retinal opsin subfamilies. This is in contrast to the 46 opsin genes characterized in *Daphnia pulex*.[47]. It remains to be seen whether these candidate gene losses are associated with the adaptation of *H. azteca* to its crepuscular visual ecology or reflect a more ancient trend in amphipods.

In contrast, the *H. azteca* genome reveals gene expansions of chemoreceptors, which may be essential for *H. azteca* given its epibenthic ecology and close association with sediments.[6] Non-insect arthropods have two major families of chemoreceptors: gustatory receptor (GR) family, an ancient lineage extending back to early animals,[49, 65-67] and the ionotropic receptors (IRs) that are a variant lineage of the ancient ionotropic glutamate receptor superfamily known only from protostomes.[49, 68] These two gene families were manually annotated in the *H. azteca* genome and improved models were generated for two other crustaceans, *D. pulex*[69, 70] and *Eurytemora affinis*[49, 71] for comparison (**S4.1**, sequences available in **S6.2**). With 155 GR genes, *H. azteca* has over twice the number of GRs compared with *D. pulex* (59)[69] and *E. affinis* (67), although many of the most recent gene duplicates are pseudogenes. Two candidate GR sugar receptors were identified in *H. azteca* and *D. pulex* (independently duplicated in both lineages), but not in *E. affinis.* Otherwise these crustacean GRs form large species-specific expanded clades with no convincing orthology with each other or other conserved insect GRs such as the Gr43a fructose receptor (**Figure S4.1.1**). *H. azteca* has 118 IR genes (2 pseudogenic) compared with updated totals of 154 in *D. pulex* (26 pseudogenic) and 22 intact genes for *E. affinis*. All three species contain single copy orthologs of the highly conserved IR genes implicated in perception of salt, amines, amino acids, humidity, and temperature in insects, including Ir25a, Ir8a (missing in *D. pulex)*, Ir76b,[72] and Ir93a.[73] The remaining divergent IRs form largely species-specific expanded lineages (**Figure S4.1.2**). The many divergent IRs and GRs in these three crustaceans presumably mediate most of their chemical sense capability, but their great divergence from the proteins of *Drosophila* for which functions are known precludes speculation as to specific roles.

Detoxification and Stress Response

Cytochrome P450s –

The cytochrome P450 superfamily of genes (P450 genes) is ubiquitous and diverse as they have been found in all domains of life and are thought to have originated over 3 billion years ago.[74] P450 genes function in metabolizing a wide range of endogenous and exogenous compounds, including toxins, drugs, plant metabolites, and signaling molecules.[75-78] In the *H. azteca* genome, we found 70 genes or gene fragments that contained a typical P450 signature (FxxGxxxC), where C is the heme thiolate ligand (**S4.3**, sequences available in **S6.3**). However, only 27 were complete genes. The 70 P450 genes were classifiable into one of four recognized P450 clans, with the CYP2 clan (**Figure S4.3.1**) being the largest with 48 genes. The most notable difference between the P450 complement of *H. azteca* relative to hexapods (insects) was the expansion of the CYP2 clan P450s. Typical of expanded clades, we found several clusters of genes (and gene fragments) of the CYP2 clan. The CYP3 and CYP4 clans in *H. azteca* were represented by eight and seven genes, respectively. The fourth P450 clan is the mitochondrial P450 clan, with at least nine genes in *H. azteca*. The number of P450s found in *H. azteca* was greater to those found in other crustaceans, including the copepods *Tigriopus japonicus* (52)[79] and *Paracyclopina nana* (46)[80] but somewhat fewer than those in hexapod taxa (including 106 in the mosquito *Anopheles gambiae* and 81 in the silkworm *Bombyx mori*). [81]

Heat shock proteins –

The heat shock protein (HSP) molecular chaperones are a highly conserved family of proteins that facilitate the refolding of denatured proteins following stress, including thermal stress, but also in response to metals and other toxicants, oxidative stress, and dehydration.[82] HSPs are divided into several families based on their molecular weight. Of the different families, HSP70, HSP90 and HSP60 play a major role in protein refolding while HSP40/J-protein is a co-factor to HSP70 and delivers nonnative proteins to HSP70.[83] HSPs were identified and annotated for each of these families (**S4.8**). The number of *hsp*70 (8 genes), *hsp*90 (3), *hsp*40 (3), and *hsp*60 (1) was well within the expected number found throughout Arthropoda.[84] Of the eight *hsp*70 genes, five were found as a gene cluster on scaffold 277, which is similar to gene clusters identified in *Drosophila melanogaster* [85] and *Aedes aegypti*.[86] In agreement with Baringou et al.,[87] the HSP70 proteins described here cannot be easily divided into inducible and cognate forms based on sequence characteristics. We instead decided to compare our eight sequences to sequence motifs described by Baringou et al.[87] and classify the *H. azteca* HSP70s according

to their framework (**Table S4.8.2**). According to these motifs and the classification methods described, all *H. azteca* sequences belong to Group A, which agrees with Baringou et al.[87] finding that all amphipod HSP70s characterized to date are Group A proteins. One HSP70 contained slightly different motif characteristics and was grouped with A4 proteins, while the remaining sequences were grouped together in A5.

Metallothionein genes –

Metallothioneins (MTs) are a group of conserved metalloproteins with a high capacity for binding metal ions. These proteins are characterized by their low molecular weight (< 10 KDa), cysteine rich composition (often over 30%), lack of secondary structure in the absence of bound metal ions, and a two domain structure dictated by the bound ions. Although their diversity makes it difficult to assign a specific function by class of MTs, their ability to bind metal ions has provided MTs with a role in detoxification, binding, and sequestration of toxic metals.[88] Four *mt* genes were identified in the *H. azteca* genome by mapping Cd responsive contigs with homology to *Callinectes sapidus* CdMT-1 (AAF08964) to the HAZT_1.0 assembly (**S4.11**). These four genes were arranged as repeats on scaffold 460 and each contained three exons, the typical gene structure of *mt*s (**Figure S4.11.1**). *Mt*-b and *mt*-d produce identical proteins of 61 amino acids, while *mt*-c is missing the downstream splice site on exon 1 and produces a truncated protein of 53 amino acids. *Mt*-a lacks a viable start codon, making it a likely pseudogene. Due to the similarity in the sequences of the remaining three genes, it is not possible to determine if they are all transcribed or regulated differently based solely on the RNAseq mapped reads. However, given their high degree of similarity and arrangement on scaffold 460, these genes are likely the result of recent gene duplications, which may provide an evolutionary advantage against high metal exposure. 1-4 *mt* genes have been identified in at least 35 other Malacostracan species. However, in most cases, the multiple MTs are not identical in amino acid sequence. For example, in the blue crab *C. sapidus* there are three MT genes. Two encode for Cd inducible forms with 76% sequence identity, while a third, codes for a longer copper-inducible form.[89] Given that our strategy for identifying the MT genes in *H. azteca* relied on using the Cd inducible gene expression set, it is possible that a fourth, copper inducible form also exists in its genome.

Ion transport proteins –

In arthropods, a subset of ion transporters are integral in maintaining cellular homeostasis and regulating epithelial transport of common ions such as H^+, Na^+, K^+, and Cl^-[90-92] [93] and are likely involved in the toxicity and uptake of metal ions. The proton pump V-type H^+ ATPase (VHA, ATP6) is an evolutionarily conserved molecular machine having a wide range of functions. VHA actively translocates H^+ across the membranes of cells and organelles allowing it to generate electrochemical H^+ gradients[94] that drive H^+-coupled substrate transport of common bioavailable cations (Na^+, K^+, Li^+).[95] VHA is a large, two domain protein complex (V1 and V0) comprised of 13 subunits, which are ubiquitous in eukaryotes and thought to be expressed in virtually every eukaryotic cell.[96] These 13 VHA subunits were identified in the *H. azteca* genome, but 2 accessory subunits were not (**S4.10**). A previous comparative analysis identified a wide range of VHA genes in the genomes of *D. melanogaster* (33), human (24), mouse (24), *C. elegans* (19), *Arabidopsis* (28), and *Saccharomyces* (15).[97] The high number of VHA genes identified in those organisms are in stark contrast to the only 13 VHA subunit genes present in *H. azteca.*

The sodium/hydrogen antiporters (NHA, SLC9B2, CPA2) are a subfamily of transmembrane ion transporters, which was only recently discovered in animal genomes and characterization in mosquito larvae.[98-100] In both arthropods and mammals, evidence indicates that NHA is coupled to VHA as a secondary electrogenic transporter for ion uptake against concentration gradients.[99, 101-104] The presence of four NHA genes in the *H. azteca* genome was unexpected, as only two NHA paralogs per genome had been found previously in animal genomes (**S4.10**).[98]

The minimal set of VHA subunits and expansion of NHA genes may have toxicological significance, particularly with respect to metal toxicity and transport. Metal speciation, and therefore toxicity, is highly pH-dependent.[105] As a regulator of pH at the epithelial membrane, and thus electrochemical transmembrane gradients, VHA may play an important role in metal uptake, ion speciation, and solubility.[106, 107] [108] The specific substrates for NHA have not yet been fully characterized for most species, but sodium and chloride are likely candidates.[109] The presence of NHA gene duplicates in the *H. azteca* genome suggest a likely adaptation for ion uptake against transmembrane concentration gradients, but may also influence metal uptake and bioavailability . As many toxic metal ions are transported across the cell membrane *via* ion channels and/or ion transporters (e.g. ATPases),[110, 111] additional transporters should be explored for their direct roles in metal uptake and indirect influence on metal toxicity.

The Toxico-responsive Genome

Contaminants such as heavy metals, organic compounds, and nanoparticles can adversely affect ecologically relevant organisms. In *H. azteca*, heavy metal contaminants (Zn, Cd) have been shown to have negative effects on the development of population growth rate, longevity, and reproduction[112-114] while metal-based nanomaterials (i.e. ZnO NMs) cause increased toxicity which may be related to enhanced bioavailability.[44, 115] Organic compounds such as pyrethroid insecticides and polychlorinated biphenyls (PCBs) can cause detrimental effects on behavior, reproduction, and development.[116-118]

A primary goal of the *H. azteca* genome project was to expand the functional gene annotations for transcripts that respond to toxicant stress, referred to as the toxicogenome. We utilized two published gene expression studies (Zn, ZnO, NPs[44], cyfluthrin[30]) as well as two unpublished gene expression sets for Cd and PCB exposure (**Table 1**). During our original investigation, we were only able to annotate a small fraction of the differentially expressed transcripts (Cd: 13%; Zn: 29%; PCB126: 0%; Cylfuthrin: 20%) (**Figure 3A**). The ability to align these transcripts to the *H. azteca* genome allowed us to identify full length transcripts and more completely assemble transcripts aligning to the same genic region of the genome in a way that was not possible with a de novo transcriptome assembler (i.e. Newbler, Roche). This increased our ability to predict gene function increasing the fraction of annotated transcripts by 10-32% (**Figure 3A**). However, we also note that for each of these chemical challenges, over half of the genes are still without annotations (Figure 3A), implying that these genes may be lineage specific. This is similar to the finding within the *D. pulex* genome that lineage specific genes were more likely to be differentially expressed following environmental challenges.[47]

Cadmium –

To explore the gene expression response and further annotate genes involved in heavy metal exposure, we conducted a gene expression study at ecologically relevant concentrations of Cd. Compared to controls, 116 genes were up-regulated in expression and 9 were down-regulated by Cd. These genes are related to several cell processes including digestion, oxygen transport, cuticular metabolism, immune function, acid-base balance, visual-sensory perception and signal transduction (**Table S5.1**). Categorizing the genes by biological processes illustrated that the metabolism of *H. azteca* was very broadly affected with the cellular metabolic processes representing the largest GO term (**Figure 3B**). Heat shock proteins (general stress response) were significantly upregulated in response to Cd (see **Table S4.8.1**) consistent with other amphipod studies[119, 120] and showing a similar response to other stressors including heat stress, oxidative stress and changes in pH.[121, 122] In addition, expression of the newly described MT

genes were also significantly induced over 15-fold by Cd (**Figure S4.11.3**). Cd exposure also induced expression of genes involved in oxidative stress including glutathione-*S*-transferase, a commonly used biomarker in toxicity tests of pollutant exposure and oxidative stress, and thioredoxin peroxidase, a gene involved in protection against reactive oxygen species (ROS). Finally, genes involved in regulation of the cuticle were also upregulated. Differential expression of chitinase and other cuticular proteins has been demonstrated previously in crustaceans in response to stress and has been correlated with impacts to growth and reproduction.[123, 124]

Zinc –

We utilized a data set originally published in Poynton et al.[44] that compared the toxicity of ZnO NPs to zinc sulfate ($ZnSO_4$) to increase the number of annotated genes that were responsive to metal exposure. Of the 60 differentially expressed genes, we annotated 25, including 15 genes that had not been annotated in the original publication (**Table S5.2**, **Figure 3A**). For example, chorion peroxidase (contig18799 in Poynton et al.[38]) was induced in both the $ZnSO_4$ and ZnO NP exposures and acts as an indicator of oxidative stress, as the gene is involved in ROS damage repair. Contig000192 in Poynton et al.[44] is another previously uncharacterized gene that was annotated as asparaginyl beta-hydroxylase-like protein, a regulator of muscle contraction and relaxation; its dysregulation suggests negative impacts to swimming behavior and movement. With the additional annotation results we were able to perform gene ontology analysis (**Table S5.2**) and observed that most genes were mapped to GO:0042221, response to chemical (**Figure 3C**).

Cyfluthrin –

The pyrethroid insecticide cyfluthrin is one the most widely applied insecticides worldwide[125, 126] and has been shown to be highly toxic to *H. azteca* (EC_{50} <1 ng/L).[116] We previously showed that cyfluthrin exposure at 1 ng/L caused differential expression of 127 sequences.[30] Through the reanalysis of this data set, we were able to annotate 33 genes and successfully mapped them to GO terms (**Table S5.3**). Many affected genes were consistent with the known mechanism of pyrethroid toxicity, showing involvement in neurological system processes, synapse organization, and transmission of nerve impulses, but also stress response such as oxidization processes, damage repair, maintaining of homeostasis, and immune response (**Figure 3D**).

PCB126 –

For *H. azteca*, PCBs represent a major exposure and accumulation threat due to their habitat and feeding behavior in benthic areas. In fishes, dioxin-like PCBs (e.g., PCB126) are highly toxic as they bind to the aryl hydrocarbon receptor and induce the expression of CYP1 genes.[127] Much less is known about this mechanism in crustaceans,[128] but in general they appear more tolerant of PCBs. Following exposure of *H. azteca* to PCB126, the most potent and ubiquitous of the PCB congeners,[129] we identified 21 differentially expressed sequences, representing seven genes, of which five were annotated (**Table S5.4**). Three of the five characterized genes are transmembrane proteins, while two are involved in endocrine processes (growth hormone, thyroid hormone). Neuroendocrine disruption of PCBs was described in crustaceans previously (see review in[130]); however, most investigations on neuroendocrine disruption to date focus on effects in vertebrates. Our study demonstrates that potential impacts of neuroendocrine disruption on invertebrates deserves further attention.

In summary, with a total of 19,936 genes including 911 manually curated genes, the genome of the *Hyalella azteca* US Lab strain provides a foundational tool for understanding the molecular ecology of benthic invertebrates as well as the mechanisms of toxicity of sediment associated pollutants. The critical gene families annotated here will serve as basis for studying toxicologically conserved pathways in invertebrates and developing adverse outcome pathways for sediment dwelling organisms. Overall, our results illustrate the advantage of applying a genome assembly to ecotoxicogenomic studies including the improved ability to annotate genes of interest and we strongly encourage the expansion of genomic, not just transcriptomic, resources for other species of ecotoxicological relevance.

The ever-growing list of chemical contaminants entering the environment poses a significant challenge in terms of risk assessment. The low-throughput and high cost of traditional toxicity testing, suggests that the need for alternative means to assess risk.[23] The characterization of ‘omics responses has emerged as a potential alternative.[25] Measures on the cellular level provide valuable information on the mode of action of uncharacterized chemicals, the health status of exposed organisms, and can act as a means to extrapolate beyond model organisms, and can be integrated into predictive risk models. The interpretation of these omics responses within the context of the well-defined *H. azteca* genome described herein will greatly expand the utility and applications of omics responses to sediment ecotoxicology and risk assessment.

Acknowledgements:

We thank the staff at the Baylor College of Medicine Human Genome Sequencing Center for their contributions, Terence Murphy at NCBI for assisting in screening for bacterial contamination in the assembly and with removing contaminants from the genome, and Dawn Oliver at Purdue University for assistance with the distribution map (Figure 1). Additional thanks to Mitch Kostich of the US EPA for his thoughtful comments on the manuscript. Funding for genome sequencing, assembly and automated annotation was provided by NHGRI grant U54 HG003273 to RAG. Funding for transcriptome sequencing was provided by NSF grant CBET-1437409 to HCP. Additional funding was provided by the State and Federal Contractors Water Agency (contract no. 15-14) to Richard Connon, UC Davis (supported SH), University of Cincinnati Faculty Development Research grant and partial support from NSF grant DEB165441 to JBB, and NSF grant DEB1257053 to JHW. MMT's work with Apollo was supported by NIH grants 5R01GM080203 from NIGMS, and 5R01HG004483 from NHGRI, and by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Supporting Information:

Additional method details (S1), detailed annotation reports (S2-S4), gene expression data tables and figures (S5), supplemental sequence files (S6) and detailed author contributions (S7) are provided in the supporting information files.

Figures and Tables:

Table 1: embedded in methods

Figure 1: Geographical distribution of seven of the most well characterized *Hyalella azteca* species groups in the United States and Canada. The two species that have been taxonomically described include *Hyalella spinicauda* and *Hyalella wellborni.*[131] The distribution shown here was described in several publications for *H. wellborni*[19, 29, 132] with additional collections by R. Cothran and G. Wellborn; and *H. spinicauda*.[19, 29, 133] The remaining five species have species level divergence in the cytochrome oxidase-I gene, but without taxonomic descriptions, they are named with the designations applied in publications describing the

collections. These include Clade 8 (also referred to as US Lab strain[29], Species C[22], OK-L[133]) with additional collections shown here by G. Wellborn, R. Cothran, M. Worsham, A. Kuzmic; Clade 1[19, 29, 132, 134, 135]; Clade 5[19]; Species B[22, 30, 133]; and Species D.[22, 30, 136] The genome sequence described here represents Clade 8 or the US Lab strain.[29] The other commonly used laboratory strain, primarily from Canada, belongs to Clade 1.[29]

Figure 2: Summary of the quality and completeness of the two *H. azteca* genome assemblies. (A) Comparison of the contig, scaffold, and assembly size between the two assemblies. (B) Comparison of predicted gene sets from the two assemblies. The original MAKER gene set for Hazt_1.0 is Hazt_0.5.3 while the gene set developed by NCBI for Hazt_2.0 is referred to as Hazt_2.0. (C) BUSCO analysis compared to other genomes of ecotoxicological relevance including the amphipod *Parhyale hawaiensis,*[48] the copepod *Eurytemora affinis,*[49] *Daphnia pulex,*[47] the aquatic midge *Chironomus tentans,*[50] the terrestrial springtail *Folsomia candida,*[53] the fathead minnow *Pimephales promelas,*[51] and the killifish *Fundulus heteroclitus*.[52] Dotted line corresponds to the total number of BUSCOs (single copy, duplicated, or fragmented) for the Hazt_2.0 assembly. (D) BUSCO comparison for the predicted gene sets. Dotted line corresponds to the total number of BUSCOs (single copy, duplicated, or fragmented) for the Hazt_2.0 gene set. (E) Percentage of RNAseq reads that mapped to the genome (left) and the predicted protein coding genes (right). Illumina data sets were acquitted from the NCBI Bioproject (PRJNA312414). Reads were mapped according to methods described in Rosendale et al.[137] and Schoville et al.[138]

Figure 3: Annotation of toxicant responsive genes. (A) The annotation of differentially expressed transcripts was significantly improved using the *H. azteca* genome. The total number of unique differentially expressed transcripts is listed below each treatment. Pie graphs represent differentially expressed contigs and illustrate the percentage of original annotations (blue) and the additional annotations that were added when contigs were aligned to the genome (orange). For PCB126, none of the contigs were annotated prior to alignment to the genome. In many cases more than one contig aligned to the same transcript; therefore, the total number of contigs is greater than the number of transcripts. (B-D) Biological processes gene ontology (GO) terms representing the differentially expressed transcripts from Cd (B), Zn and ZnO NPs (C), and cyfluthrin (D). The number of genes mapped to each of the GO terms is shown by the length of the bars, while the percentage of total transcripts is marked at the end of each bar.

For PCB126, none of the 12 annotated transcripts were mapped to biological processes GO terms. Similar graphs for molecular function GO terms can be found in the supporting information (**Figure S5**).

Figure 1:

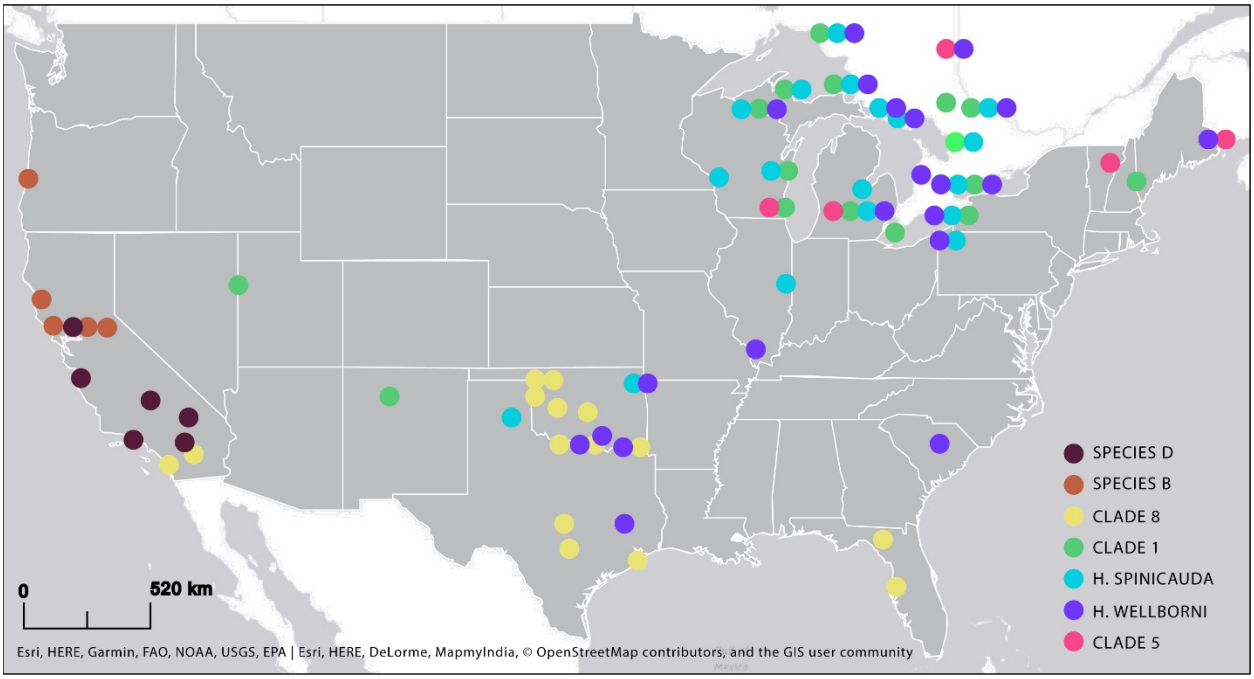

SPECIES D
SPECIES B
CLADE 8
CLADE 1
H. SPINICAUDA
H. WELLBORNI
CLADE 5
0
520 km
Esri, HERE, Garmin, FAO, NOAA, USGS, EPA | Esri, HERE, DeLorme, MapmyIndia, © OpenStreetMap contributors, and the GIS user community

Figure 2:

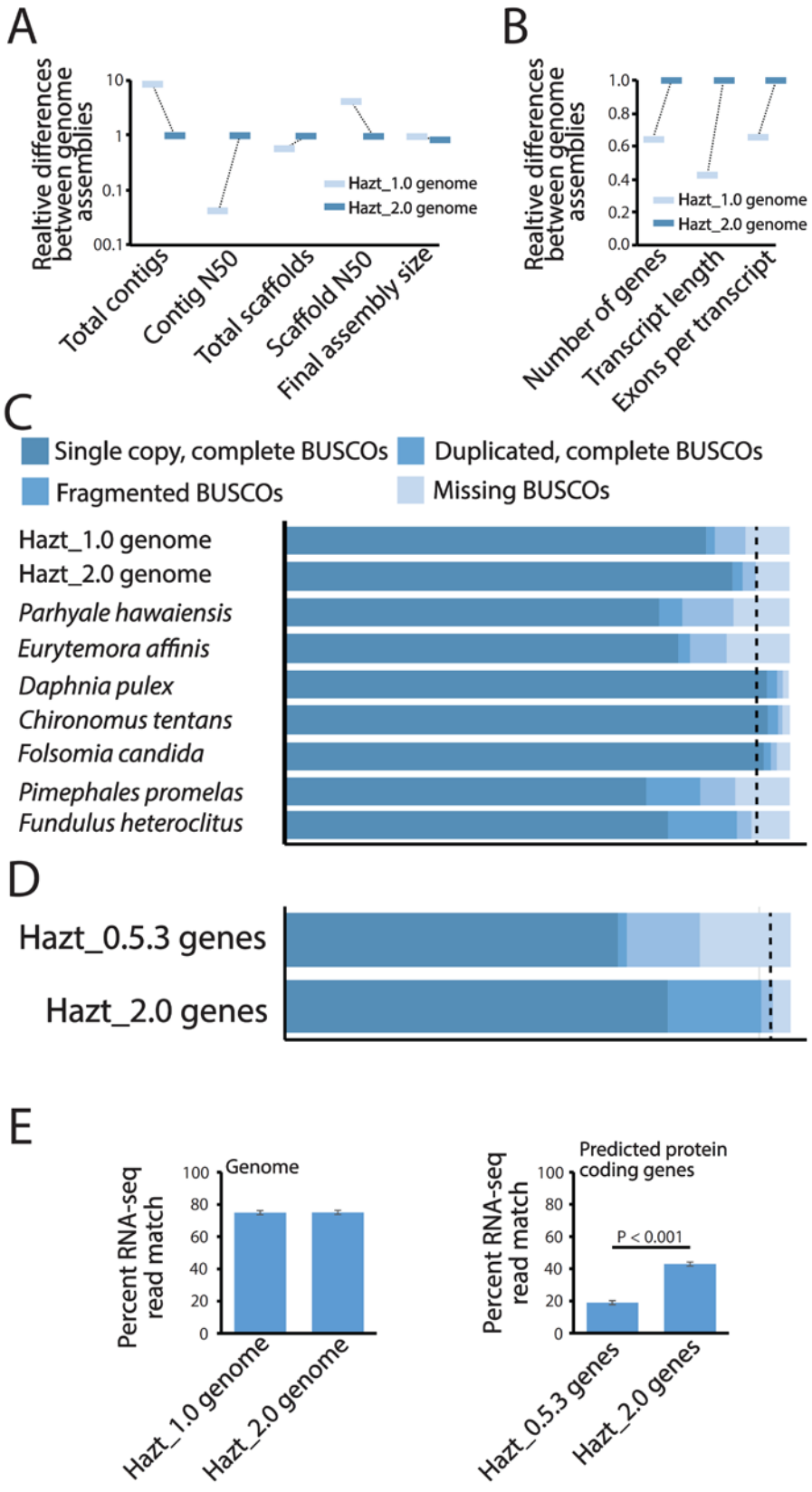
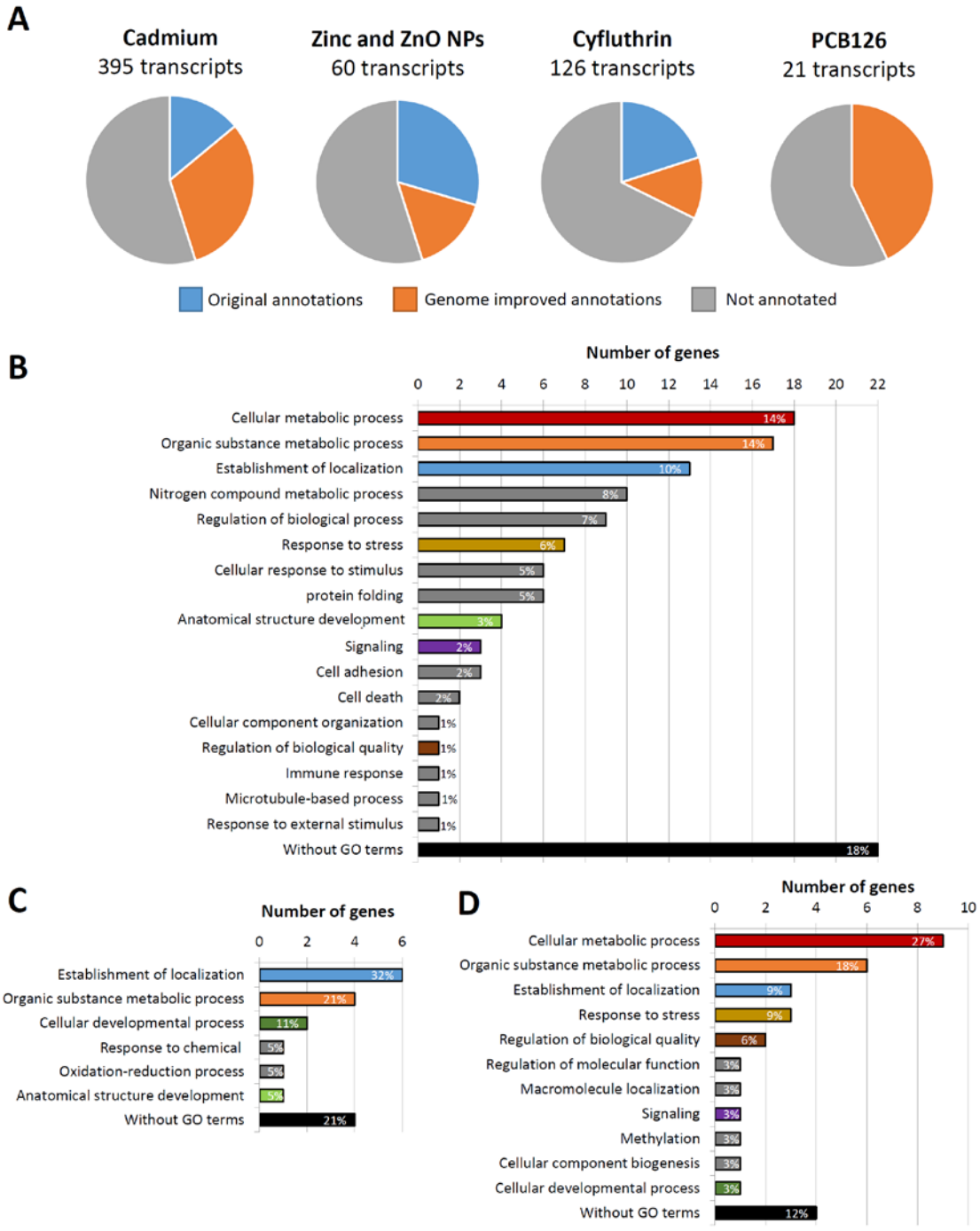

A
Realtive differences between genome assemblies
10
1
0.1
00.1
Hazt_1.0 genome
Hazt_2.0 genome
Total contigs
Contig N50
Total scaffolds
Scaffold N50
Final assembly size
B
Realtive differences between genome assemblies
1.0
0.8
0.6
0.4
0.2
0.0
Hazt_1.0 genome
Hazt_2.0 genome
Number of genes
Transcript length
Exons per transcript
C
Single copy, complete BUSCOs
Duplicated, complete BUSCOs
Fragmented BUSCOs
Missing BUSCOs
Hazt_1.0 genome
Hazt_2.0 genome
Parhyale hawaiensis
Eurytemora affinis
Daphnia pulex
Chironomus tentans
Folsomia candida
Pimephales promelas
Fundulus heteroclitus
D
Hazt_0.5.3 genes
Hazt_2.0 genes
E
Percent RNA-seq read match
Genome
100
80
60
40
20
0
Hazt_1.0 genome
Hazt_2.0 genome
Predicted protein coding genes
P < 0.001
Hazt_0.5.3 genes
Hazt_2.0 genes

Figure 3:


A
Cadmium
395 transcripts
Zinc and ZnO NPs
60 transcripts
Cyfluthrin
126 transcripts
PCB126
21 transcripts
Original annotations
Genome improved annotations
Not annotated
B
Number of genes
0 2 4 6 8 10 12 14 16 18 20 22
Cellular metabolic process 14%
Organic substance metabolic process 14%
Establishment of localization 10%
Nitrogen compound metabolic process 8%
Regulation of biological process 7%
Response to stress 6%
Cellular response to stimulus 5%
protein folding 5%
Anatomical structure development 3%
Signaling 2%
Cell adhesion 2%
Cell death 2%
Cellular component organization 1%
Regulation of biological quality 1%
Immune response 1%
Microtubule-based process 1%
Response to external stimulus 1%
Without GO terms 18%
C
Number of genes
0 2 4 6
Establishment of localization 32%
Organic substance metabolic process 21%
Cellular developmental process 11%
Response to chemical 5%
Oxidation-reduction process 5%
Anatomical structure development 5%
Without GO terms 21%
D
Number of genes
0 2 4 6 8 10
Cellular metabolic process 27%
Organic substance metabolic process 18%
Establishment of localization 9%
Response to stress 9%
Regulation of biological quality 6%
Regulation of molecular function 3%
Macromolecule localization 3%
Signaling 3%
Methylation 3%
Cellular component biogenesis 3%
Cellular developmental process 3%
Without GO terms 12%

References:

(1) Chapman, P. M., Current approaches to developing sediment quality criteria. *Environ. Toxicol. Chem.* **1989,** *8*, (7), 589-599.

(2) US EPA. *The Incidence and Severity of Sediment Contamination in Surface Waters of the United States*; EPA-823-R-04-007; Office of Science and Technology, US Environmental Protection Agency: Washington D.C., November, 2004.

(3) Gavrilescu, M.; Demnerová, K.; Aamand, J.; Agathos, S.; Fava, F., Emerging pollutants in the environment: present and future challenges in biomonitoring, ecological risks and bioremediation. *New Biotechnol.* **2015,** *32*, (1), 147-156.

(4) Stone, W. W.; Gilliom, R. J.; Ryberg, K. R., Pesticides in US streams and rivers: occurrence and trends during 1992–2011. In ACS Publications: 2014.

(5) Lee, K. E.; Langer, S. K.; Menheer, M. A.; Foreman, W. T.; Furlong, E. T.; Smith, S. G. *Chemicals of emerging concern in water and bottom sediment in Great Lakes areas of concern, 2010 to 2011-Collection methods, analyses methods, quality assurance, and data*; 723; Reston, VA, 2012; p 36.

(6) Wang, F.; Goulet, R. R.; Chapman, P. M., Testing sediment biological effects with the freshwater amphipod *Hyalella azteca*: the gap between laboratory and nature. *Chemosphere* **2004,** *57*, 1713-1724.

(7) Norberg-King, T. J.; Sibley, P. K.; Burton, G. A.; Ingersoll, C. G.; Kemble, N. E.; Ireland, S.; Mount, D. R.; Rowland, C. D., Interlaboratory evaluation of Hyalella azteca and Chironomus tentans short-term and long-term sediment toxicity tests. *Environ. Toxicol. Chem.* **2006,** *25*, (10), 2662-2674.

(8) Dussault, E. B.; Balakrishnan, V. K.; Sverko, E.; Solomon, K. R.; Sibley, P. K., Toxicity of human pharmaceuticals and personal care products to benthic invertebrates. *Environ. Toxicol. Chem.* **2008,** *27*, (2), 425-432.

(9) You, J.; Pehkonen, S.; Weston, D. P.; Lydy, M. J., Chemical availability and sediment toxicity of pyrethroid insecticides to *Hyalella azteca*: Application to field sediment with unexpectedly low toxicity. *Environ. Toxicol. Chem.* **2008,** *27*, (10), 2124-2130.

(10) U.S. EPA *Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates*; US Environmental Protection Agency: 2000.

(11) Taylor, L. N.; Novak, L.; Rendas, M.; Antunes, P. M. C.; Scroggins, R. P., Validation of a new standardized test method for the freshwater amphipod Hyalella azteca: Determining the chronic effects of silver in sediment. *Environ. Toxicol. Chem.* **2016,** *35*, (10), 2430-2438.

(12) Strong, D. R., Life history variation among populations of an Amphipod (*Hyalella azteca*). *Ecology* **1972,** *53*, (6), 1103-1111.

(13) Wellborn, G. A., The Mechanistic Basis of Body-Size Differences between 2 Hyalella (Amphipoda) Species. *J. Freshwater Ecol.* **1994,** *9*, (2), 159-168.

(14) Wellborn, G. A., Size-Biased Predation and Prey Life-Histories - a Comparative-Study of Fresh-Water Amphipod Populations. *Ecology* **1994,** *75*, (7), 2104-2117.

(15) Wellborn, G. A., Predator Community Composition and Patterns of Variation in Life-History and Morphology among Hyalella (Amphipoda) Populations in Southeast Michigan. *Am. Midl. Nat.* **1995,** *133*, (2), 322-332.

(16) Duan, Y. H.; Guttman, S. I.; Oris, J. T., Genetic differentiation among laboratory populations of *Hyalella azteca*: Implications for toxicology. *Environ. Toxicol. Chem.* **1997,** *16*, (4), 691-695.

(17) Thomas, P. E.; Blinn, D. W.; Keim, P., Genetic and behavioural divergence among desert spring amphipod populations. *Freshwater Biol.* **1997,** *38*, (1), 137-143.

(18) Hogg, I. D.; Larose, C.; de Lafontaine, Y.; Doe, K. G., Genetic evidence for a Hyalella species complex within the Great Lakes St Lawrence River drainage basin: implications for ecotoxicology and conservation biology. *Can. J. Zool.* **1998,** *76*, (6), 1134-1140.

(19) Witt, J. D.; Hebert, P. D., Cryptic species diversity and evolution in the amphipod genus Hyalella within central glaciated North America: a molecular phylogenetic approach. *Can. J. Fish. Aquat. Sci.* **2000,** *57*, (4), 687-698.
(20) Duan, Y. H.; Guttman, S. I.; Oris, J. T.; Bailer, A. J., Genotype and toxicity relationships among Hyalella azteca: I. Acute exposure to metals or low pH. *Environ. Toxicol. Chem.* **2000,** *19*, (5), 1414-1421.
(21) Wellborn, G. A.; Cothran, R.; Bartholf, S., Life history and allozyme diversification in regional ecomorphs of the *Hyalella azteca* (Crustacea : Amphipoda) species complex. *Biol. J. Linn. Soc.* **2005,** *84*, (2), 161-175.
(22) Major, K. M.; Weston, D. P.; Lydy, M. J.; Wellborn, G. A.; Poynton, H. C., Unintentional exposure to terrestrial pesticides drives widespread and predictable evolution of resistance in freshwater crustaceans. *Evol. Appl.* **2017,** (in press), 1-14. https://doi.org/10.1111/eva.12584.
(23) Nationa Research Council, *Toxicity testing in the 21st century: a vision and a strategy*. National Academies Press: 2007.
(24) 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), 730-41.
(25) Villeneuve, D. L.; Garcia-Reyero, N., Vision & strategy: Predictive ecotoxicology in the 21st century. *Environ. Toxicol. Chem.* **2011,** *30*, (1), 1-8.
(26) Goldstone, J.; Hamdoun, A.; Cole, B.; Howard-Ashby, M.; Nebert, D.; Scally, M.; Dean, M.; Epel, D.; Hahn, M.; Stegeman, J., The chemical defensome: environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. *Dev. Biol.* **2006,** *300*, (1), 366-384.
(27) Perkins, E. J.; Ankley, G. T.; Crofton, K. M.; Garcia-Reyero, N.; LaLone, C. A.; Johnson, M. S.; Tietge, J. E.; Villeneuve, D. L., Current perspectives on the use of alternative species in human health and ecological hazard assessments. *Environ. Health Perspect.* **2013,** *121*, (9), 1002.
(28) LaLone, C. A.; Villeneuve, D. L.; Burgoon, L. D.; Russom, C. L.; Helgen, H. W.; Berninger, J. P.; Tietge, J. E.; Severson, M. N.; Cavallin, J. E.; Ankley, G. T., Molecular target sequence similarity as a basis for species extrapolation to assess the ecological risk of chemicals with known modes of action. *Aquat. Toxicol.* **2013,** *144*, 141-154.
(29) Major, K.; Soucek, D. J.; Giordano, R.; Wetzel, M. J.; Soto-Adames, F., The common ecotoxicology laboratory strain of *Hyalella azteca* is genetically distinct from most wild strains sampled in Eastern North America. *Environ. Toxicol. Chem.* **2013,** *32*, (11), 2637-47.
(30) Weston, D. P.; Poynton, H. C.; Wellborn, G. A.; Lydy, M. J.; Blalock, B. J.; Sepulveda, M. S.; Colbourne, J. K., Multiple origins of pyrethroid insecticide resistance across the species complex of a nontarget aquatic crustacean, *Hyalella azteca*. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, (41), 16532-7.
(31) Powell, J. R.; Evans, B. R., How Much Does Inbreeding Reduce Heterozygosity? Empirical Results from *Aedes aegypti*. *Am. J. Trop. Med. Hyg.* **2017,** *96*, (1), 157-158.
(32) Gnerre, S.; MacCallum, I.; Przybylski, D.; Ribeiro, F. J.; Burton, J. N.; Walker, B. J.; Sharpe, T.; Hall, G.; Shea, T. P.; Sykes, S., High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108*, (4), 1513-1518.
(33) Pryszcz, L. P.; Gabaldón, T., Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **2016,** *44*, (12), e113-e113.
(34) Cantarel, B. L.; Korf, I.; Robb, S. M.; Parra, G.; Ross, E.; Moore, B.; Holt, C.; Alvarado, A. S.; Yandell, M., MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **2008,** *18*, (1), 188-196.

(35) Hughes, D. S. T.; Richards, S.; Poynton, H. *Hyalella azteca* Genome Annotations v0.5.3. Ag Data Commons, 2018, http://dx.doi.org/10.15482/USDA.ADC/1415993.
(36) Thibaud-Nissen, F.; Souvorov, A.; Murphy, T.; DiCuccio, M.; Kitts, P., Eukaryotic Genome Annotation Pipeline. In *The NCBI Handbook* 2nd ed.; National Center for Biotechnology Information (US): Bethesda, MD, 2013.
(37) Yandell, M.; Ence, D., A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* **2012,** *13*, (5), 329.
(38) Lee, E.; Helt, G. A.; Reese, J. T.; Munoz-Torres, M. C.; Childers, C. P.; Buels, R. M.; Stein, L.; Holmes, I. H.; Elsik, C. G.; Lewis, S. E., Web Apollo: a web-based genomic annotation editing platform. *Genome Biol.* **2013,** *14*, (8), R93.
(39) Poynton, H.; Gibbs, R. A.; Worley, K. C.; Murali, S. C.; Lee, S. L.; Muzny, D. M.; Hughes, D. S. T.; Chao, H.; Dinh, H.; Doddapaneni, H.; Qu, J.; Dugan, S.; Han, Y.; Richards, S. *Hyalella azteca* Genome Assembly 1.0. Ag Data Commons, 2018, http://dx.doi.org/10.15482/USDA.ADC/1415994.
(40) Skinner, M. E.; Uzilov, A. V.; Stein, L. D.; Mungall, C. J.; Holmes, I. H., JBrowse: a next-generation genome browser. *Genome Res.* **2009,** *19*, (9), 1630-1638.
(41) Poynton, H.; Bain, P.; Benoit, J.; Boucher, P.; Chen, M.-J. M.; Chen, S.; Christie, A. E.; Connolly, J.; Cridge, A.; Etxabe, A.; Glazier, A.; Hasenbein, S.; Hughes, D.; Jennings, E.; Jones, J.; Kearns, P.; Lambert, F.; Lin, Y.-Y.; Major, C.; Manny, A.; Mathers, N.; Mechler-Hickson, A.; Munoz-Torres, M.; Poelchau, M.; Robertson, H.; Scanlan, J.; Szuter, E.; Werren, J.; Wood, S. A.; Winchell, K.; Richards, S. *Hyalella azteca* Official Gene Set v1.0. Ag Data Commons, 2018, http://dx.doi.org/10.15482/USDA.ADC/1415995.
(42) Poelchau, M.; Childers, C.; Moore, G.; Tsavatapalli, V.; Evans, J.; Lee, C.-Y.; Lin, H.; Lin, J.-W.; Hackett, K., The i5k Workspace@ NAL—enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Res.* **2014,** *43*, (D1), D714-D719.
(43) Christie, A. E.; Cieslak, M. C.; Roncalli, V.; Lenz, P. H.; Major, K. M.; Poynton, H. C., Prediction of a peptidome for the ecotoxicological model Hyalella azteca (Crustacea; Amphipoda) using a de novo assembled transcriptome. *Mar. Genom.* **2018,** *38,* 67-88.
(44) Poynton, H. C.; Lazorchak, J. M.; Impellitteri, C. A.; Blalock, B.; Smith, M. E.; Struewing, K.; Unrine, J.; Roose, D., Toxicity and transcriptomic analysis in *Hyalella azteca* suggests increased exposure and susceptibility of epibenthic organisms to zinc oxide nanoparticles. *Environ. Sci. Technol.* **2013,** *47*, (16), 9453-60.
(45) Vergilino, R.; Dionne, K.; Nozais, C.; Dufresne, F.; Belzile, C., Genome size differences in Hyalella cryptic species. *Genome* **2012,** *55*, (2), 134-139.
(46) Simão, F. A.; Waterhouse, R. M.; Ioannidis, P.; Kriventseva, E. V.; Zdobnov, E. M., BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015,** *31*, (19), 3210-3212.
(47) 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.; 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.; Tsuchiya, D.; Tu, H.; Vos, H.; Wang, M.; Wolf, Y. I.; Yamagata, H.; Yamada, T.; Ye, Y.; Shaw, J. R.; Andrews, J.; Crease, T. J.; Tang, H.; 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), 555-61.
(48) Kao, D.; Lai, A. G.; Stamataki, E.; Rosic, S.; Konstantinides, N.; Jarvis, E.; Di Donfrancesco, A.; Pouchkina-Stancheva, N.; Semon, M.; Grillo, M., The genome of the

crustacean *Parhyale hawaiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *eLife* **2016,** *5*, e20062.
(49) Eyun, S.-i.; Young Soh, H.; Posavi, M.; Munro, J. B.; Hughes, D. S.; Murali, S. C.; Qu, J.; Dugan, S.; Lee, S. L.; Chao, H., Evolutionary history of chemosensory-related gene families across the Arthropoda. *Mol. Biol. Evol.* **2017**, msx147.
(50) Kutsenko, A.; Svensson, T.; Nystedt, B.; Lundeberg, J.; Björk, P.; Sonnhammer, E.; Giacomello, S.; Visa, N.; Wieslander, L., The Chironomus tentans genome sequence and the organization of the Balbiani ring genes. *BMC Genomics* **2014,** *15*, (1), 819.
(51) Burns, F. R.; Cogburn, A. L.; Ankley, G. T.; Villeneuve, D. L.; Waits, E.; Chang, Y. J.; Llaca, V.; Deschamps, S. D.; Jackson, R. E.; Hoke, R. A., Sequencing and de novo draft assemblies of a fathead minnow (Pimephales promelas) reference genome. *Environ. Toxicol. Chem.* **2016,** *35*, (1), 212-217.
(52) Reid, N. M.; Jackson, C. E.; Gilbert, D.; Minx, P.; Montague, M. J.; Hampton, T. H.; Helfrich, L. W.; King, B. L.; Nacci, D. E.; Aluru, N., The landscape of extreme genomic variation in the highly adaptable Atlantic killifish. *Genome Biol. Evol.* **2017,** *9*, (3), 00-00.
(53) Faddeeva-Vakhrusheva, A.; Kraaijeveld, K.; Derks, M. F.; Anvar, S. Y.; Agamennone, V.; Suring, W.; Kampfraath, A. A.; Ellers, J.; Le Ngoc, G.; Gestel, C. A., Coping with living in the soil: the genome of the parthenogenetic springtail Folsomia candida. *BMC Genomics* **2017,** *18*, (1), 493.
(54) Wheeler, D.; Redding, A. J.; Werren, J. H., Characterization of an ancient lepidopteran lateral gene transfer. *PloS One* **2013,** *8*, (3), e59262.
(55) Alneberg, J.; Bjarnason, B. S.; de Bruijn, I.; Schirmer, M.; Quick, J.; Ijaz, U. Z.; Lahti, L.; Loman, N. J.; Andersson, A. F.; Quince, C., Binning metagenomic contigs by coverage and composition. *Nat. Methods* **2014,** *11*, (11), 1144-1146.
(56) Yoshida, S.; Hiraga, K.; Takehana, T.; Taniguchi, I.; Yamaji, H.; Maeda, Y.; Toyohara, K.; Miyamoto, K.; Kimura, Y.; Oda, K., A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* **2016,** *351*, (6278), 1196-1199.
(57) Bombelli, P.; Howe, C. J.; Bertocchini, F., Polyethylene bio-degradation by caterpillars of the wax moth *Galleria mellonella*. *Curr. Biol. 27*, (8), R292-R293.
(58) Glastad, K.; Hunt, B.; Yi, S.; Goodisman, M., DNA methylation in insects: on the brink of the epigenomic era. *Insect Mol. Biol.* **2011,** *20*, (5), 553-565.
(59) Hunt, B. G.; Glastad, K. M.; Yi, S. V.; Goodisman, M. A., The function of intragenic DNA methylation: insights from insect epigenomes. In Oxford University Press: 2013.
(60) Glastad, K. M.; Gokhale, K.; Liebig, J.; Goodisman, M. A., The caste-and sex-specific DNA methylome of the termite Zootermopsis nevadensis. *Scientific Reports* **2016,** *6*, 37110.
(61) Choudhuri, S., Small noncoding RNAs: biogenesis, function, and emerging significance in toxicology. *J. Biochem. Mol. Toxic.* **2010,** *24*, (3), 195-216.
(62) Chen, S.; McKinney, G. J.; Nichols, K. M.; Colbourne, J. K.; Sepúlveda, M. S., Novel cadmium responsive microRNAs in Daphnia pulex. *Environ. Sci. Technol.* **2015,** *49*, (24), 14605-14613.
(63) Chen, S.; Nichols, K. M.; Poynton, H. C.; Sepúlveda, M. S., MicroRNAs are involved in cadmium tolerance in Daphnia pulex. *Aquat. Toxicol.* **2016,** *175*, 241-248.
(64) Carlini, D. B.; Satish, S.; Fong, D. W., Parallel reduction in expression, but no loss of functional constraint, in two opsin paralogs within cave populations of Gammarus minus (Crustacea: Amphipoda). *BMC Evol. Biol.* **2013,** *13*, (1), 89.
(65) Eyun, S.-i.; Soh, H. Y.; Posavi, M.; Munro, J. B.; Hughes, D. S. T.; Murali, S. C.; Qu, J.; Dugan, S.; Lee, S. L.; Chao, H.; Dinh, H.; Han, Y.; Doddapaneni, H.; Worley, K. C.; Muzny, D. M.; Park, E.-O.; Silva, J. C.; Gibbs, R. A.; Richards, S.; Lee, C. E., Evolutionary History of Chemosensory-Related Gene Families across the Arthropoda. *Mol. Biol. Evol.* **2017,** *34*, (8), 1838-1862.

(66) Saina, M.; Busengdal, H.; Sinigaglia, C.; Petrone, L.; Oliveri, P.; Rentzsch, F.; Benton, R., A cnidarian homologue of an insect gustatory receptor functions in developmental body patterning. *Nat. Commun.* **2015,** *6: 6243*, 1-12.
(67) Robertson, H. M., The insect chemoreceptor superfamily is ancient in animals. *Chem. Senses* **2015,** *40*, (9), 609-614.
(68) Rytz, R.; Croset, V.; Benton, R., Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in Drosophila and beyond. *Insect Biochem. Molec.* **2013,** *43*, (9), 888-897.
(69) Peñalva-Arana, D. C.; Lynch, M.; Robertson, H. M., The chemoreceptor genes of the waterflea Daphnia pulex: many Grs but no Ors. *BMC Evol. Biol.* **2009,** *9*, (1), 79.
(70) Croset, V.; Rytz, R.; Cummins, S. F.; Budd, A.; Brawand, D.; Kaessmann, H.; Gibson, T. J.; Benton, R., Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* **2010,** *6*, (8), e1001064.
(71) Robertson, H. M., Noncanonical GA and GG 5′ Intron Donor Splice Sites Are Common in the Copepod Eurytemora affinis. *G3-Genes Genom. Genet.* **2017,** *7*, (12), 3967-3969.
(72) Croset, V.; Schleyer, M.; Arguello, J. R.; Gerber, B.; Benton, R., A molecular and neuronal basis for amino acid sensing in the Drosophila larva. *Scientific Reports* **2016,** *6*, 34871.
(73) Knecht, Z. A.; Silbering, A. F.; Cruz, J.; Yang, L.; Croset, V.; Benton, R.; Garrity, P. A., Ionotropic Receptor-dependent moist and dry cells control hygrosensation in Drosophila. *eLife* **2017,** *6*, e26654.
(74) Danielson, P., The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr. Drug Metab.* **2002,** *3*, (6), 561-597.
(75) Miller, W. L.; Chung, B.-c., The First Defect in Electron Transfer to Mitochondrial P450 Enzymes. In Oxford University Press: 2016.
(76) Morale, M. C.; L'Episcopo, F.; Tirolo, C.; Giaquinta, G.; Caniglia, S.; Testa, N.; Arcieri, P.; Serra, P.-A.; Lupo, G.; Alberghina, M., Loss of aromatase cytochrome P450 function as a risk factor for Parkinson's disease? *Brain Res. Rev.* **2008,** *57*, (2), 431-443.
(77) Sigel, A.; Sigel, H.; Sigel, R. K., *The ubiquitous roles of cytochrome P450 proteins*. Wiley Onlie Library: 2007; Vol. 3.
(78) Wahlang, B.; Falkner, K. C.; Cave, M. C.; Prough, R. A., Role of Cytochrome P450 Monooxygenase in Carcinogen and Chemotherapeutic Drug Metabolism. *Adv. Pharmacol.* **2015,** *74*, 1-33.
(79) Han, J.; Won, E.-J.; Hwang, D.-S.; Shin, K.-H.; Lee, Y. S.; Leung, K. M.-Y.; Lee, S.-J.; Lee, J.-S., Crude oil exposure results in oxidative stress-mediated dysfunctional development and reproduction in the copepod Tigriopus japonicus and modulates expression of cytochrome P450 (CYP) genes. *Aquat. Toxicol.* **2014,** *152*, 308-317.
(80) Han, J.; Won, E.-J.; Kim, H.-S.; Nelson, D. R.; Lee, S.-J.; Park, H. G.; Lee, J.-S., Identification of the full 46 cytochrome P450 (CYP) complement and modulation of CYP expression in response to water-accommodated fractions of crude oil in the cyclopoid copepod Paracyclopina nana. *Environ. Sci. Technol.* **2015,** *49*, (11), 6982-6992.
(81) Feyereisen, R., Evolution of insect P450. *Biochem. Soc. T.* **2006,** *34*, (6), 1252-1255.
(82) Zhao, L.; Jones, W., Expression of heat shock protein genes in insect stress responses. *Invertebrate Surviv. J.* **2012,** *90*, 93-101.
(83) Richter, K.; Haslbeck, M.; Buchner, J., The heat shock response: life on the verge of death. *Molecul. Cell* **2010,** *40*, (2), 253-266.
(84) Nguyen, A. D.; Gotelli, N. J.; Cahan, S. H., The evolution of heat shock protein sequences, cis-regulatory elements, and expression profiles in the eusocial Hymenoptera. *BMC Evol. Biol.* **2016,** *16*, (1), 15.
(85) Bettencourt, B. R.; Feder, M. E., Hsp70 duplication in the Drosophila melanogaster species group: how and when did two become five? *Mol. Biol. Evol.* **2001,** *18*, (7), 1272-1282.
(86) Gross, T. L.; Myles, K. M.; Adelman, Z. N., Identification and characterization of heat shock 70 genes in *Aedes aegypti* (Diptera: Culicidae). *J. Med. Entomol.* **2014,** *46*, (3), 496-504.

(87) Baringou, S.; Rouault, J.-D.; Koken, M.; Hardivillier, Y.; Hurtado, L.; Leignel, V., Diversity of cytosolic HSP70 Heat Shock Protein from decapods and their phylogenetic placement within Arthropoda. *Gene* **2016,** *591*, (1), 97-107.
(88) Amiard, J.-C.; Amiard-Triquet, C.; Barka, S.; Pellerin, J.; Rainbow, P., Metallothioneins in aquatic invertebrates: their role in metal detoxification and their use as biomarkers. *Aquat. Toxicol.* **2006,** *76*, (2), 160-202.
(89) Syring, R. A.; Brouwer, T. H.; Brouwer, M., Cloning and sequencing of cDNAs encoding for a novel copper-specific metallothionein and two cadmium-inducible metallothioneins from the blue crab Callinectes sapidus. *Comp. Biochem. Phys. C* **2000,** *125*, (3), 325-332.
(90) Wieczorek, H.; Beyenbach, K. W.; Huss, M.; Vitavska, O., Vacuolar-type proton pumps in insect epithelia. *J. Exp. Biol.* **2009,** *212*, (11), 1611-1619.
(91) Lee, C. E.; Kiergaard, M.; Gelembiuk, G. W.; Eads, B. D.; Posavi, M., Pumping ions: rapid parallel evolution of ionic regulation following habitat invasions. *Evolution* **2011,** *65*, (8), 2229-2244.
(92) McNamara, J. C.; Faria, S. C., Evolution of osmoregulatory patterns and gill ion transport mechanisms in the decapod Crustacea: a review. *J. Comp. Physiol. B* **2012,** *182*, (8), 997-1014.
(93) Lee, C. E., Evolutionary mechanisms of habitat invasions, using the copepod *Eurytemora affinis* as a model system. *Evol. Appl.* **2016,** *9*, (1), 248-270.
(94) Wieczorek, H.; Putzenlechner, M.; Zeiske, W.; Klein, U., A vacuolar-type proton pump energizes K+/H+ antiport in an animal plasma membrane. *J. Biol. Chem.* **1991,** *266*, (23), 15340-15347.
(95) Maxson, M. E.; Grinstein, S., The vacuolar-type H+-ATPase at a glance–more than a proton pump. In The Company of Biologists Ltd: 2014.
(96) Beyenbach, K. W.; Wieczorek, H., The V-type H+ ATPase: molecular structure and function, physiological roles and regulation. *J. Exp. Biol.* **2006,** *209*, (4), 577-589.
(97) Allan, A. K.; Du, J.; Davies, S. A.; Dow, J. A., Genome-wide survey of V-ATPase genes in Drosophila reveals a conserved renal phenotype for lethal alleles. *Physiol. Genomics* **2005,** *22*, (2), 128-138.
(98) Brett, C. L.; Donowitz, M.; Rao, R., Evolutionary origins of eukaryotic sodium/proton exchangers. *Am. J. Physiol.-Cell Ph.* **2005,** *288*, (2), C223-C239.
(99) Rheault, M. R.; Okech, B. A.; Keen, S. B.; Miller, M. M.; Meleshkevitch, E. A.; Linser, P. J.; Boudko, D. Y.; Harvey, W. R., Molecular cloning, phylogeny and localization of AgNHA1: the first Na+/H+ antiporter (NHA) from a metazoan, *Anopheles gambiae*. *J. Exp. Biol.* **2007,** *210*, (21), 3848-3861.
(100) Donowitz, M.; Tse, C. M.; Fuster, D., SLC9/NHE gene family, a plasma membrane and organellar family of Na+/H+ exchangers. *Mol. Aspects Med.* **2013,** *34*, (2), 236-251.
(101) Harvey, W. R., Voltage coupling of primary H+ V-ATPases to secondary Na+-or K+-dependent transporters. *J. Exp. Biol.* **2009,** *212*, (11), 1620-1629.
(102) Xiang, M. A.; Linser, P. J.; Price, D. A.; Harvey, W. R., Localization of two Na+-or K+-H+ antiporters, AgNHA1 and AgNHA2, in Anopheles gambiae larval Malpighian tubules and the functional expression of AgNHA2 in yeast. *J. Insect Physiol.* **2012,** *58*, (4), 570-579.
(103) Kondapalli, K. C.; Kallay, L. M.; Muszelik, M.; Rao, R., Unconventional chemiosmotic coupling of NHA2, a mammalian Na+/H+ antiporter, to a plasma membrane H+ gradient. *J. Biol. Chem.* **2012,** *287*, (43), 36239-36250.
(104) Battaglino, R. A.; Pham, L.; Morse, L. R.; Vokes, M.; Sharma, A.; Odgren, P. R.; Yang, M.; Sasaki, H.; Stashenko, P., NHA-oc/NHA2: a mitochondrial cation–proton antiporter selectively expressed in osteoclasts. *Bone* **2008,** *42*, (1), 180-192.
(105) Schubauer-Berigan, M. K.; Dierkes, J. R.; Monson, P. D.; Ankley, G. T., pH-Dependent toxicity of Cd, Cu, Ni, Pb and Zn to *Ceriodaphnia dubia*, *Pimephales promelas*, *Hyalella azteca* and *Lumbriculus variegatus*. *Environ. Toxicol. Chem.* **1993,** *12*, (7), 1261-1266.

(106) Jackson, B.; Lasier, P.; Miller, W.; Winger, P., Effects of calcium, magnesium, and sodium on alleviating cadmium toxicity to *Hyalella azteca*. *B. Environ. Contam. Tox.* **2000,** *64*, (2), 279-286.
(107) Wang, Z.; Meador, J. P.; Leung, K. M., Metal toxicity to freshwater organisms as a function of pH: A meta-analysis. *Chemosphere* **2016,** *144*, 1544-1552.
(108) de Schamphelaere, K. A.; Janssen, C. R., A biotic ligand model predicting acute copper toxicity for Daphnia magna: the effects of calcium, magnesium, sodium, potassium, and pH. *Environ. Sci. Technol.* **2002,** *36*, (1), 48-54.
(109) Chintapalli, V. R.; Kato, A.; Henderson, L.; Hirata, T.; Woods, D. J.; Overend, G.; Davies, S. A.; Romero, M. F.; Dow, J. A., Transport proteins NHA1 and NHA2 are essential for survival, but have distinct transport modalities. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112*, (37), 11720-11725.
(110) Foulkes, E. C., Transport of Toxic Heavy Metals Across Cell Membranes (44486). *P. Soc. Exp. Biol. Med.* **2000,** *223*, (3), 234-240.
(111) Argüello, J. M.; Raimunda, D.; González-Guerrero, M., Metal Transport across Biomembranes: Emerging Models for a Distinct Chemistry. *J. Biol. Chem.* **2012,** *287*, (17), 13510-13517.
(112) Liber, K.; Doig, L. E.; White-Sobey, S. L., Toxicity of uranium, molybdenum, nickel, and arsenic to Hyalella azteca and Chironomus dilutus in water-only and spiked-sediment toxicity tests. *Ecotox. Environ. Safe.* **2011,** *74*, (5), 1171-1179.
(113) Nguyen, L. T. H.; Muyssen, B. T. A.; Janssen, C. R., Single versus combined exposure of Hyalella azteca to zinc contaminated sediment and food. *Chemosphere* **2012,** *87*, (1), 84-90.
(114) Ball, A. L.; Borgmann, U.; Dixon, D. G., Toxicity of a cadmium-contaminated diet to Hyalella azteca. *Environ. Toxicol. Chem.* **2006,** *25*, (9), 2526-2532.
(115) Wallis, L. K.; Diamond, S. A.; Ma, H.; Hoff, D. J.; Al-Abed, S. R.; Li, S., Chronic TiO2 nanoparticle exposure to a benthic organism, Hyalella azteca: impact of solar UV radiation and material surface coatings on toxicity. *Sci. Total Environ.* **2014,** *499*, (Supplement C), 356-362.
(116) Hasenbein, S.; Connon, R. E.; Lawler, S. P.; Geist, J., A comparison of the sublethal and lethal toxicity of four pesticides in Hyalella azteca and Chironomus dilutus. *Environ. Sci. Pollut. Res.* **2015,** *22*, (15), 11327-11339.
(117) Hasenbein, S.; Lawler, S. P.; Geist, J.; Connon, R. E., A long-term assessment of pesticide mixture effects on aquatic invertebrate communities. *Environ. Toxicol. Chem.* **2016,** *35*, (1), 218-232.
(118) Ingersoll, C. G.; Brunson, E. L.; Dwyer, F. J.; Hardesty, D. K.; Kemble, N. E., Use of sublethal endpoints in sediment toxicity tests with the amphipod Hyalella azteca. *Environ. Toxicol. Chem.* **1998,** *17*, (8), 1508-1523.
(119) Hook, S. E.; Twine, N. A.; Simpson, S. L.; Spadaro, D. A.; Moncuquet, P.; Wilkins, M. R., 454 pyrosequencing-based analysis of gene expression profiles in the amphipod Melita plumulosa: Transcriptome assembly and toxicant induced changes. *Aquat. Toxicol.* **2014,** *153*, (Sp. Iss. SI), 73-88.
(120) Werner, I.; Nagel, R., Stress proteins HSP60 and HSP70 in three species of amphipods exposed to cadmium, diazinon, dieldrin and fluoranthene. *Environ. Toxicol. Chem.* **1997,** *16*, (11), 2393-2403.
(121) Liang, P.; MacRae, T. H., Molecular chaperones and the cytoskeleton. *J. Cell Science* **1997,** *110*, (13), 1431-1440.
(122) Watanabe, M.; Suzuki, T., Cadmium-induced synthesis of HSP70 and a role of glutathione in Euglena gracilis. *Redox Report* **2004,** *9*, (6), 349-353.
(123) Poynton, H. C.; Varshavsky, J. R.; Chang, B.; Cavigiolio, G.; Chan, S.; Holman, P. S.; Loguinov, A. V.; Bauer, D. J.; Komachi, K.; Theil, E. C.; Perkins, E. J.; Hughes, O.; Vulpe, C. D., Daphnia magna Ecotoxicogenomics Provides Mechanistic Insights into Metal Toxicity. *Environ. Sci. Technol.* **2007,** *41*, (3), 1044-1050.

(124) Poynton, H. C.; Loguinov, A. V.; Varshavsky, J. R.; Chan, S.; Perkins, E. J.; Vulpe, C. D., Gene expression profiling in Daphnia magna part I: concentration-dependent profiles provide support for the no observed transcriptional effect level. *Environ. Sci. Technol.* **2008,** *42*, (16), 6250-6256.
(125) Spurlock, F.; Lee, M., Synthetic Pyrethroid Use Patterns, Properties, and Environmental Effects. *In Synthetic Pyrethroids. ACS Symposium Series*. **2008,** *991*, 3-25.
(126) Stehle, S.; Schulz, R., Agricultural insecticides threaten surface waters at the global scale. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112*, (18), 5750-5755.
(127) Nebert, D. W., Aryl hydrocarbon receptor (AHR):"pioneer member" of the basic-helix/loop/helix per-Arnt-sim (bHLH/PAS) family of "sensors" of foreign and endogenous signals. *Prog. Lipid Res.* **2017,** *67*, 38-57.
(128) Kim, B.-M.; Rhee, J.-S.; Hwang, U.-K.; Seo, J. S.; Shin, K.-H.; Lee, J.-S., Dose- and time-dependent expression of aryl hydrocarbon receptor (AhR) and aryl hydrocarbon receptor nuclear translocator (ARNT) in PCB-, B[a]P-, and TBT-exposed intertidal copepod *Tigriopus japonicus*. *Chemosphere* **2015,** *120*, 398-406.
(129) Lee, H. G.; Yang, J. H., PCB126 induces apoptosis of chondrocytes via ROS-dependent pathways. *Osteoarthr. Cartilage* **2012,** *20*, (10), 1179-1185.
(130) Waye, A.; Trudeau, V. L., Neuroendocrine Disruption: More than Hormones are Upset. *J. Toxicol. Env. Heal. B.* **2011,** *14*, (5-7), 270-291.
(131) Soucek, D. J.; Lazo-Wasem, E. A.; Taylor, C. A.; Major, K. M., Description of Two New Species of Hyalella (Amphipoda: Hyalellidae) from Eastern North America with a Revised Key to North American Members of the Genus. *J. Crustacean Biol.* **2015,** *35*, (6), 814-829.
(132) Dionne, K.; Vergilino, R.; Dufresne, F.; Charles, F.; Nozais, C., No Evidence for Temporal Variation in a Cryptic Species Community of Freshwater Amphipods of the Hyalella azteca Species Complex. *Diversity* **2011,** *3*, (3), 390.
(133) Wellborn, G. A.; Broughton, R. E., Diversification on an ecologically constrained adaptive landscape. *Mol. Ecol.* **2008,** *17*, (12), 2927-2936.
(134) Galbreath, J. G. S.; Smith, J. E.; Becnel, J. J.; Butlin, R. K.; Dunn, A. M., Reduction in post-invasion genetic diversity in Crangonyx pseudogracilis (Amphipoda: Crustacea): a genetic bottleneck or the work of hitchhiking vertically transmitted microparasites? *Biol. Invasions* **2010,** *12*, (1), 191-209.
(135) Baird, D. J.; Pascoe, T. J.; Zhou, X.; Hajibabaei, M., Building freshwater macroinvertebrate DNA-barcode libraries from reference collection material: formalin preservation vs specimen age. *J. N. Am. Benthol. Soc.* **2011,** *30*, (1), 125-130.
(136) Witt, J. D.; Threloff, D. L.; Hebert, P. D., DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Mol. Ecol.* **2006,** *15*, (10), 3073-3082.
(137) Rosendale, A. J.; Romick-Rosendale, L. E.; Watanabe, M.; Dunlevy, M. E.; Benoit, J. B., Mechanistic underpinnings of dehydration stress in the American dog tick revealed through RNA-Seq and metabolomics. *J. Exp. Biol.* **2016,** *219*, (12), 1808-1819.
(138) Schoville, S. D.; Chen, Y. H.; Andersson, M. N.; Benoit, J. B.; Bhandari, A.; Bowsher, J. H.; Brevik, K.; Cappelle, K.; Chen, M.-J. M.; Childers, A. K., A model species for agricultural pest genomics: the genome of the Colorado potato beetle, Leptinotarsa decemlineata (Coleoptera: Chrysomelidae). *bioRxiv* **2017**, 192641.