Comparative Biochemistry and Physiology, Part D 20 (2016) 27–40

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part D

journal homepage: www.elsevier.com/locate/cbpd

Transcriptome discovery in non-model wild fish species for the development of quantitative transcript abundance assays

Cassidy M. Hahn [a,b,*], Luke R. Iwanowicz [b], Robert S. Cornman [b], Patricia M. Mazik [a], Vicki S. Blazer [b]

[a] West Virginia University, School of Natural Resources, 322 Percival Hall, Morgantown, WV 26506, USA
[b] US Geological Survey, Leetown Science Center, 11649 Leetown Road, Kearneysville, WV 25430, USA

ARTICLE INFO

Article history:
Received 21 January 2016
Received in revised form 4 June 2016
Accepted 8 July 2016
Available online 12 July 2016

ABSTRACT

Environmental studies increasingly identify the presence of both contaminants of emerging concern (CECs) and legacy contaminants in aquatic environments; however, the biological effects of these compounds on resident fishes remain largely unknown. High throughput methodologies were employed to establish partial transcriptomes for three wild-caught, non-model fish species; smallmouth bass (*Micropterus dolomieu*), white sucker (*Catostomus commersonii*) and brown bullhead (*Ameiurus nebulosus*). Sequences from these transcriptome databases were utilized in the development of a custom nCounter CodeSet that allowed for direct multiplexed measurement of 50 transcript abundance endpoints in liver tissue. Sequence information was also utilized in the development of quantitative real-time PCR (qPCR) primers. Cross-species hybridization allowed the smallmouth bass nCounter CodeSet to be used for quantitative transcript abundance analysis of an additional non-model species, largemouth bass (*Micropterus salmoides*). We validated the nCounter analysis data system with qPCR for a subset of genes and confirmed concordant results. Changes in transcript abundance biomarkers between sexes and seasons were evaluated to provide baseline data on transcript modulation for each species of interest.

Published by Elsevier Inc.

1. Introduction

Chemicals of emerging concern (CECs) are increasingly detected in aquatic environments worldwide (Birch et al., 2015; Blair et al., 2013; Lee et al., 2012; Li et al., 2010; Lopez et al., 2015; Nilsen et al., 2014; Stuart et al., 2012). These contaminants can be defined as synthetic or naturally occurring chemicals that are newly recognized, not commonly monitored, but have the potential to enter the environment and cause known or suspected adverse ecological and/or human health effects. This group includes current use pesticides, personal care products, pharmaceuticals, detergents, plasticizers, brominated flame retardants, and biogenic hormones, both human and animal (Klaper and Welch, 2011; Klecka et al., 2010). These compounds are defined not only for their potential detrimental effects on the environment, but by the lack of regulation pertaining to their release into air, water and sediment. In most cases, CECs are present as complex mixtures which may also include legacy contaminants such as heavy metals, polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs). Despite widespread detection in aquatic environments, little information is available regarding the biological effects of exposure to these complex mixtures.

Resident fish species are ideal bioindicators and integrators of chemical mixture effects as they are continually exposed to CECs and legacy contaminants and serve as sentinels for the evaluation of the impacts in aquatic ecosystems. Contaminant monitoring studies have identified indications of reproductive endocrine disruption, such as intersex (testicular oocytes), in the pelagic smallmouth bass *Micropterus dolomieu* and largemouth bass *M. salmoides* (Blazer et al., 2007; Blazer et al., 2012; Hinck et al., 2009; Iwanowicz et al., 2009; Sepulveda et al., 2002; Tetreault et al., 2011). Benthic species, including the brown bullhead *Ameiurus nebulosus* and white sucker *Catostomus commersonii*, have been widely used in ecotoxicological studies evaluating the presence of liver and skin tumors and other adverse effects (Baumann et al., 1996; Blazer et al., 2014; Blazer et al., 2009a, 2009b; Hayes et al., 1990; Pinkney et al., 2011; Premdas et al., 1995; Pyron et al., 2001; Rafferty et al., 2009; Smith et al., 1989), immune dysfunction (Iwanowicz et al., 2012) and reproductive endocrine disruption (McMaster, 2001; Woodling et al., 2006). These types of organismal level effects are often indicative of long term exposure to contaminants; therefore, bioindicators at the molecular level, which assess current and/or recent environmental conditions, can be incorporated to improve upon these assessments (Garcia-Reyero et al., 2008; Garcia-Reyero et al., 2014). The linkage of molecular initiation events

* Corresponding author at: West Virginia University, School of Natural Resources, Morgantown, WV 26506, USA.
E-mail address: cmhahn@usgs.gov (C.M. Hahn).

http://dx.doi.org/10.1016/j.cbd.2016.07.001
1744-117X/Published by Elsevier Inc.

to adverse outcomes at higher levels of biological organization produces information that can be used for risk assessment by ecotoxicologists.

Unfortunately, much of the genetic information available for fish species is from model organisms such as the fathead minnow *Pimephales promelas*, zebrafish *Danio rerio* and medaka *Oryzias latipes*. While this information has greatly contributed to our understanding of molecular mechanisms, similar information for wild fish species commonly used in environmental studies is lacking. Hence, our objective was the development of a method to measure transcript abundance in non-model, but environmentally relevant, species that could be utilized in conjunction with higher-order organismal health indicators. High throughput sequencing technology was employed to establish partial transcriptomes for smallmouth bass, white sucker and brown bullhead. These nucleotide databases were then used in the development of custom nCounter analysis assays for the evaluation of differential transcript abundance in liver tissue of wild-caught fishes. Endpoints to be evaluated during assay development were determined by literature review of commonly modulated genes in model species exposed to CECs and legacy contaminants. We present baseline data on differences in transcript abundance between sexes, seasons and species.

2. Methods

2.1. Transcriptome sequencing

The initial databases for brown bullhead (BBH) and smallmouth bass (SMB) were created using wild caught individuals from multiple geographic areas and laboratory fish that had been exposed to 17β-estradiol in a previous study (Robertson et al., 2009). Liver, gonad, anterior kidney and spleen tissue were pooled. Total RNA was extracted using an E.N.Z.A. ® Total RNA Kit (Omega Biotek, Norcross, GA). Total RNA was sent to Cofactor Genomics (St. Louis, MO) for library construction and sequencing on an Illumnia GAIIx (San Diego, CA). Additional sequencing was performed on a 454 GS FLX instrument (Roche, Branford, CT) at Duke Institute for Genome Science and Policy (Durham, NC). For each species, Illumina 60-bp paired-end reads were quality and adapter trimmed with CLC Genomics Workbench (v5). Reads were then assembled *de novo* with that software using three different kmer lengths (35, 43, 51) and other settings at default values. These three initial assemblies were merged with a secondary assembly using automatic kmer selection and other default parameters. The 454 runs were assembled with MIRA3 (Chevreux et al., 2004) using the following quick switches: "denovo, est., normal, 454". Finally, the SMB 454 and Illumina contigs were merged using CAP3 (Huang and Madan, 1999) with an overlap length of 60 and percent identity of 90, whereas the BBH 454 and Illumina contigs were merged with CLC Genomics (v5) using default de novo assembly parameters.

White sucker (WHS) transcriptome sequencing was performed using tissue composites from male and female wild fish collected from the Great Lakes region (Hahn et al., 2015). Total RNA was isolated from skin and liver samples using an E.N.Z.A. ® RNA Kit (Omega Biotek, Norcross, GA) from nine individuals, pooled and enriched for non-ribosomal RNA using a RiboZero rRNA removal kit (Epicentre, Madison, WI). The resulting RNA preparation was sequenced as 100-bp paired-end reads on an Illumina HiSeq2000 (San Diego, CA) operated by the Institute for Genome Sciences (Baltimore, MD). Additional sequencing was performed on a 454 GS Junior (Roche, Branford, CT). Illumina sequences were then assembled *de novo* into contigs using CLC workbench (v6) using the multi-kmer approach as above, but with five values (21, 31, 41, 51 and 61). The 454 reads were also assembled with CLC Genomics (v6) using automatic kmer selection and default parameters. The final merging of these six initial assemblies was also performed with CLC Genomics (v6) using automatic kmer selection and default parameters. These assembly approaches were selected from a variety of iterations performed with multiple programs, and were chosen based on assembly N50, number of Blastx hits to the proteome of *Danio rerio* (version Zv9; accession GCF_000002035.4), and low apparent redundancy (proportion of unique kmers of length 21).

The Blast2GO plugin (v2.8) for CLC Genomics Workbench (v8) was used for functional annotation of sequence data. First, sequences were used as queries in Blastx searches against the uniprot databases for medaka and zebrafish applying the BLOSUM62 matrix and setting the lower threshold at an e-value of $1 \times e^{-6}$. We retained the top 10 alignments for each sequence. Next, functional information for all blast hits was retrieved from the Gene Ontology database and annotated using default parameters in Blast2GO. GO-Slim reduction was performed using default parameters in Blast2GO in order to restrict ontology content to broad functional categories.

2.2. Wild fish sampling

We attempted to collect 20 mature (defined as > 250 mm in length) fish of each species at each sampling site and season (Table 1). Fish were collected using a variety of collection methods including electrofishing, fyke nets and trap nets. Each species was collected during both the fall (late September to early October) and the spring (mid-April to early May) from sites within the Great Lakes region (Fig. 1). Smallmouth bass (SMB) were collected from the St. Louis River (Duluth, MN) in the fall of 2010 (n = 7) and spring of 2011 (n = 6). Largemouth bass (LMB) were collected from the Ashtabula River (Ashtabula, OH) in the fall (n = 20) and spring of 2011 (n = 20). White sucker (WHS) were collected from Swan Creek, a tributary of the Maumee River (Toledo, OH) in the fall of 2010 (n = 18) and spring of 2011 (n = 20). Brown bullhead (BBH) were collected in the fall (n = 20) and spring (n = 20) of 2011 from Conneaut Creek (Conneaut, OH). Fish were anesthetized in a holding tank that contained a lethal dose of tricaine methanesulfonate (MS-222, Argent Labs, Redmond, WA). A comprehensive necropsy-based assessment was conducted on each fish collected. Small sections of liver were collected and preserved in RNA later® (ThermoFisher, Waltham, MA) for transcript abundance analysis.

2.3. Transcript abundance analysis

Transcriptome shotgun assembly sequence databases (SMB: SRX156704, SRX199239; LMB: SRX1436666, SRX1339139; BBH: SRX199312, SRX148685) were searched by Blastx to identify *a priori* genes of interest. Genes of interest included those suspected or known to be modulated in response to CEC exposure including endocrine disrupting compounds, heavy metals, dioxin-like molecules and other environmental stressors. In addition to host genes this method was also able to identify the presence of pathogens that may

Table 1
River, season, species and sex of wild fish collected in 2010 and 2011 for which transcript abundance was determined.

Site	Date	Brown bullhead	White sucker	Largemouth bass	Smallmouth bass
St. Louis River (Duluth, MN)	Fall 2010				3 F 4 M
	Spring 2011				3 F 3 M
Swan Creek (Toledo, OH)	Fall 2010		6 F 12 M		
	Spring 2011		10 F 10 M		
Ashtabula River (Ashtabula, OH)	Spring 2011			8 F 12 M	
	Fall 2011			10 F 10 M	
Conneaut Creek (Conneaut, OH)	Spring 2011	9 F 11 M			
	Fall 2011	10 F 10 M			