Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

Prediction of a peptidome for the ecotoxicological model *Hyalella azteca* (Crustacea; Amphipoda) using a *de novo* assembled transcriptome

Andrew E. Christie[a,*], Matthew C. Cieslak[a], Vittoria Roncalli[a], Petra H. Lenz[a], Kaley M. Major[b], Helen C. Poynton[b,*]

[a] Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822, USA
[b] School for the Environment, University of Massachusetts Boston, 100 Morrissey Boulevard, Boston, MA 02125, USA

ARTICLE INFO

Keywords:
de novo transcriptome assembly
Bioinformatics
Endocrinology
Peptidergic signaling
Sediment ecotoxicology

ABSTRACT

Due to its sensitivity to many environmental and anthropogenic stressors, including a wide range of chemical compounds, *Hyalella azteca*, a freshwater amphipod, has emerged as one of the most commonly used invertebrates for ecotoxicological assessment.Peptidergic signaling systems are key components in the control of organism-environment interactions, and there is a growing literature suggesting that they are targets of a number of aquatic toxicants.Interestingly, and despite its model species status in the field of ecotoxicology, little is known about the peptide hormones of *H. azteca*.Here, a transcriptome was produced for this species using the *de novo* assembler Trinity and mined for sequences encoding putative peptide precursors; the transcriptome was assembled from 460,291,636 raw reads and consists of 133,486 unique transcripts.Seventy-six sequences encoding peptide pre/preprohormones were identified from this transcriptome, allowing for the prediction of 202 distinct peptides, which included members of the allatostatin A, allatostatin B, allatostatin C, allatotropin, bursicon, CCHamide, corazonin, crustacean cardioactive peptide, crustacean hyperglycemic hormone/molt-inhibiting hormone, ecdysis-triggering hormone, eclosion hormone, elevenin, FMRFamide-like peptide, glycoprotein hormone, GSEFLamide, inotocin, leucokinin, myosuppressin, neuropeptide F, orcokinin, orcomyotropin, pigment dispersing hormone, proctolin, pyrokinin, red pigment concentrating hormone, RYamide, short neuropeptide F, SIFamide, sulfakinin, tachykinin-related peptide and trissin families.These peptides expand the known peptidome for *H. azteca* approximately nine-fold, forming a strong foundation for future studies of peptidergic control, including disruption by aquatic toxicants, in this important ecotoxicological model.

1. Introduction

Sediments form the foundation of aquatic ecosystems, providing habitat for a diversity of organisms, including a wide variety of invertebrate species.Sediments are also repositories for chemical contaminants, which, in some instances, can accumulate to toxic levels; a 1998 survey of sediment samples by the US Environmental Protection Agency (USEPA) estimated there to be 1.2 billion cubic yards of contaminated sediment in aquatic habitats in the United States alone (USEPA, 1998).Because of the importance of sediments for aquatic ecosystems, understanding the toxicity of sediment-associated chemicals to the organisms that inhabit them is essential for the effective management and remediation of contaminated habitats.

Hyalella azteca is a freshwater amphipod native to North America.In the ecosystems it inhabits, *H. azteca* is a benthic species, living on or near the sediment surface where it scavenges for algae and detritus (Wang et al., 2004).This species' nearly continuous contact with sediment, in combination with its short generation time, sensitivity to many environmental and anthropogenic stressors, and ecological importance makes it a model for evaluation of sediment toxicity and the bioavailability of sediment contaminants (Ingersoll et al., 1995).For these reasons, *H. azteca* was selected by the USEPA for standardized method development, and is currently one of the most commonly used invertebrate models for toxicological assessment (Environment Canada, 2013; USEPA, 2000).The model species status of *H. azteca* puts it at the center of a number of important regulatory decisions (e.g., Clean Water Act; Toxic Substances Control Act; Federal Insecticide, Fungicide, and Rodenticide Act; Comprehensive Environmental Response, Compensation, and Liability Act), including, but not limited to, the listing of impaired sites and the registration of new chemicals and pesticides (e.g., USEPA, 2000).

Despite its designation as a model species for the field of

* Corresponding authors.
E-mail addresses: crabman@pbrc.hawaii.edu (A.E. Christie), Helen.Poynton@umb.edu (H.C. Poynton).

https://doi.org/10.1016/j.margen.2017.12.003
Received 14 November 2017; Received in revised form 29 November 2017; Accepted 5 December 2017

ecotoxicology, few studies have gone beyond traditional toxicity endpoints (e.g., determination of LD50's) to investigate the physiological mechanisms underlying sediment toxicity in *H. azteca* (e.g., Poynton et al., 2013; Weston et al., 2013), or for that matter, in any amphipod (e.g., Biales et al., 2013; Hook et al., 2014; Trapp et al., 2014).Thus, developing a mechanistic understanding of toxicity in *H. azteca* is important for several reasons.First, *H. azteca* is a species complex that has diverged across North America over the past 12 million years (Witt and Hebert, 2000), with different clades varying in their response to pollutants.Studies investigating the chemical sensitivity within this species complex have found that genetic variation correlates with differences in sensitivity (e.g., Duan et al., 2000; Hogg et al., 1998; Soucek et al., 2015).Thus, understanding the mechanisms underlying differential sensitivity to environmental toxicants is imperative for protecting local amphipod populations.In addition, new strategies for toxicity testing have been proposed to address the magnitude of chemicals produced and deficiencies in current testing approaches (e.g., National Research Council, 2007).The new testing paradigm relies on identifying key toxicity pathways at the cellular level and developing adverse outcome pathway models that connect cellular perturbations with ecologically relevant effects (Ankley et al., 2010; Villeneuve and Garcia-Reyero, 2011).Thus, to keep sediment toxicity testing current with emerging approaches in ecological health assessments, providing mechanistic data for *H. azteca* is essential.At the core of this deficiency is a lack of molecular resources (genomic, transcriptomic, proteomic and metabolomic data) for *H. azteca* (Poynton and Vulpe, 2009; Van Aggelen et al., 2010).

Hormonal signaling is a key component in the ability of an organism to adapt to changing environmental conditions.While many different types of chemicals can serve as circulating and/or locally released hormones, peptides are by far the largest and most diverse single class of these compounds (Kastin, 2006).In arthropods, endocrine disruption, including disruption of peptidergic systems, is an area of major environmental concern (e.g., Ford, 2012; Hyne, 2011; LeBlanc, 2007; Rodríguez et al., 2007; Soin and Smagghe, 2007); it is also one area of focus for the development of next generation pesticides that are more ecologically friendly (e.g., Altstein, 2001, 2004; Audsley and Down, 2015; Scherkenbeck and Zdobinsky, 2009; Van Hiel et al., 2010; Xie et al., 2015; Zhang et al., 2015a, 2015b).Interestingly, and despite their use in the fields of ecotoxicology and toxicogenomics, little work has focused on identifying the native peptide hormones present in amphipod species.In fact, there is just one prior study that focused on large-scale peptide discovery in amphipods, and here, just 23 peptides were predicted for *H. azteca* (Christie, 2014a).

The development and public deposition of transcriptomic resources for crustaceans has provided a powerful resource for peptide discovery in a diverse array of species from this arthropod subphylum (Bao et al., 2015; Christie, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2016a, 2016b; Christie and Chi, 2015a; Christie et al., 2008, 2010a, 2013, 2015, 2016, 2017a; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010; Toullec et al., 2013; Veenstra, 2015, 2016; Ventura et al., 2014; Yan et al., 2012). Via this strategy, large peptidomes have been generated for a wide variety of crustaceans (Bao et al., 2015; Christie, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2016a, 2016b; Christie and Chi, 2015a; Christie et al., 2013, 2015, 2017a; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010; Toullec et al., 2013; Veenstra, 2015, 2016; Ventura et al., 2014; Yan et al., 2012), including the amphipod *Echinogammarus veneris* (Christie, 2014a); the structures of 110 distinct peptides were recently predicted for *E. veneris* using a publicly accessible transcriptome (Christie, 2014a).In an effort to expand the peptidome for *H. azteca*, as well as to add to the molecular resources available for it in a general sense, a transcriptome was generated and mined for putative peptide-encoding transcripts.Specifically, known arthropod preprohormones, including those from *E. veneris* (Christie, 2014a), and, in several cases, ones from *H. azteca* itself (Christie, 2014a), were used to search the *de novo* assembly for sequences encoding homologous proteins.The proteins deduced from these transcripts were then used to predict the mature structures of the peptides contained within the precursors.This strategy allowed for the identification of 202 distinct mature peptides, expanding the predicted peptidome of *H. azteca* approximately nine-fold.The transcripts, precursor proteins and putative mature peptides identified here provide the first significant resource for initiating investigations of peptidergic signaling in *H. azteca*, including how peptides contribute to organism-environment interactions in this important ecotoxicological model species.

2. Materials and methods

2.1. Hyalella azteca culture

H. azteca (US Laboratory Strain; Major et al., 2013) were cultured according to standard test conditions (USEPA, 2000).In brief, ten animals aged 7–8 days were placed in 300 ml high-form beakers containing 5 ml of sand substrate and 175 ml of overlying control water (15 mg/l Cl and 0.02 mg/l Br); animals were fed a diet of 0.5 mg diatoms (*Thalassiosira weissflogii*; Reed Mariculture, Campbell, CA, USA) and 0.25 mg Tetramin fishfood (Tetra, Blacksburg, VA, USA) daily, with water renewals at least three times per week.After ten days, animals were harvested from the beakers with a pipet and placed immediately in RNAlater (Ambion, ThermoFisher Scientific, Waltham, MA, USA). Twelve biological replicates of ten animals each were generated.

In addition to the 12 samples just described, two separate collections of animals were also made from laboratory cultures and pooled.The first pool included three adult males and four adult females; all females contained embryos, and one male and one female was an actively mating pair.The second pool contained nine juveniles aged < 24-h to 3-days old.All animals were depurated in standard culture media for 4-h prior to RNA isolation.Following its isolation, RNA from these two collections was combined in equal amounts to create a pool of RNA representing "mixed life stages".

2.2. RNA isolation, cDNA library construction, and Illumina sequencing

For RNA extraction, RNAlater was removed from each *H. azteca* sample and the sample subsequently rinsed with TRI Reagent (Molecular Research Center, Cincinnati, OH, USA) to remove residual RNAlater.Following washing, 0.5 ml of fresh TRI Reagent was added to each tube and the sample was homogenized using a TissueLyser II bead mill (QIAGEN, Valencia, CA, USA). RNA was extracted according to the manufacture-supplied protocol and DNAse I treated using QIAGEN RNAeasy on-column digestion.RNA quality was assessed using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) at the University of Massachusetts Boston Center for Personalized Cancer Therapy (CPCT) Genomics Core.

400 ng of RNA was used for library preparation using an Illumina TruSeq Stranded mRNA kit (Illumina, San Diego, CA, USA) following the manufacturer's guidelines and using a different Illumina adapter for each library.Library quantity and fragment pool length was accessed using an Agilent 2100 Bioanalzyer; four samples required cleaned-up using an additional bead wash step to minimize adapter dimers.The final 13 libraries were sequenced using a paired-end (PE75) Rapid Run protocol on an Illumina HiSeq2500 instrument at the University of Massachusetts Boston CPCT Genomics Core producing approximately 35 million reads per sample (see Table 1).Sequence files were parsed using bcl2fastq (v2.17, Illumina). Raw reads were assessed for quality using FastQC software (v0.10.1, Babraham Bioinformatics, Babraham Institute, Babraham, Cambridge, United Kingdom). Low quality reads with Phred quality scores below 20 (fewer than 1%) and Illumina adapters were removed from the dataset using Trim Galore (v0.37, stringency 3, error rate 0, paired) (Babraham Bioinformatics, Babraham Institute, Babraham, Cambridge, United Kingdom) at the Massachusetts