Title: Integrated transcriptomic and metabolomic analyses predict the effects on the gonads of female sharpbelly (*Hemiculter lucidus*) exposed to municipal wastewater effluent in Ba River

Jiahua Guo[1], Yongrong Hao[1], Linrong Han[1], Jiezhang Mo[2], Hisato Iwata[3]

1. Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China.

2. Department of Chemistry, City University of Hong Kong, Kowloon, Hong Kong SAR, China

3. Center for Marine Environmental Studies, Ehime University, Bunkyo-cho 2-5, Matsuyama, Ehime prefecture, Japan. 790-8577.

1. Aims

In the present study, the influence of Municipal wastewater effluent (MWWE) on the ovary development of female sharpbelly living near wastewater treatment plant outfall in Ba River was evaluated. This would be achieved by the two following objectives: (1) to examine the ovarian histology altered by MWWE; (2) to unravel the pathways disrupted possibly by chemicals in MWWE at both transcriptional and metabolic levels, predicting the potential adverse outcome in ovaries of wild fish.

2 Materials and methods

2.1 Fish collection and tissue sampling

In Ba River, three sampling sites were chosen to evaluate the effects on the ovarian development of wild fish exposed to MWWE (Fig. 1). The site U was located approximate 5 km up-stream from the MWWTP outfall (109°01'2.56''E, 34°22'4.76''N) as a control point for this study; the site M situated near the MWWTP effluent outlet (109°01'11.78''E, 34°23'6.20''N), where the direct impact of effluents on fish can be assessed; the site D (109°01'1.01'' E, 34°24' 57.21''N) was located approximate 5 km down-stream from the site M to evaluate the effects of the diluted MWWTP discharges on wild fish. For transcriptomic and metabolomic profiling, the ovaries were quickly removed, snap frozen, and kept at −80 °C prior to the extraction of the total RNA and metabolites. In addition, the dissected ovaries were kept in the fixative at 4°C before further processing in histological analysis.

2.2 Transcriptomic analysis

2.2.1 Illumina sequencing, *de novo* assembly, and functional annotation

Then, Illumina NovaSeq was used for Next-generation sequencing by Shanghai Biozeron Biothchology Co., Ltd. (China). The RNA sequencing used paired-end method and the read length was 150 base pair.

Subsequent to obtaining the raw reads, Cutadapt (v1.16) software (Martin, 2011) was used to prune the sequencing files. The clean data from all samples were subjected to *de novo* assembly by Trinity (Grabherr et al., 2011) to identify unigenes. For the annotation of gene function, each unigene was searched against the NCBI protein non-redundant (NR), Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genome (KEGG) databases (Conesa et al., 2005; Moriya et al., 2007). The fragments Per Kilo bases per Million fragments (FPKM) were quantified from the normalization of clean reads (Li and Dewey, 2011). Principal component

analysis (PCA) was then conducted using “DESeq” package in R (Anders and Huber, 2010) to cluster the samples.

2.2.2 DEGs-based enrichment analysis

The FPKMs between different sites were analyzed using “DESeq” package to obtain *p* value. Each set of Differentially expressed genes (DEGs) was separately submitted for GO and KEGG pathway analyses, where a $p < 0.05$ was deemed as significantly enriched.

2.3 Untargeted metabolomic analysis

2.3.1 LC-MS/MS analysis

Metabolites in 50 mg of each gonad samples were extracted with 1000 μL of extract solution constituted by acetonitrile, methanol, and water at a ratio 2: 2: 1. Subsequent to spiking 1 μg mL^{-1} of 2-Chloro-L-phenylalanine as an internal standard, the homogenates were subjected to homogenization (40 Hz) for 4 min and sonication for three times. The mixtures were incubated for 1h at -20 ^{o}C prior to centrifugation at 12,000 rpm (4 ^{o}C, 15 min). An equal aliquot of the supernatants from each sample was blended and used as the quality control (QC). Several quality control measures were used to ensure the stability of the instrument, such as monitoring the base peak chromatograms (BPCs) for three QC samples and the total ion chromatograms (TICs) of internal standard 2-Chloro-L-phenylalanine for three QC samples in both positive and negative modes (Fig. S1-2). Besides, the principal component analysis (PCA) of metabolites and the correlation of metabolites for QC samples were illustrated in Fig. S3-4. The variation in metabolites of three QC samples in PCA were less than two standard deviations and the correlation coefficients were above 0.8, suggesting the reliability of metabolomics results. Based on these results, it was believed that the metabolites were quantified with confidence.

The LC-MS/MS analyses were performed using an UHPLC system (1290, Agilent Technologies) with a UPLC HSS T3 column (2.1 mm × 100 mm, 1.8 μm) coupled to Q Exactive mass spectrometer (Orbitrap MS, Thermo). Details on the instrumental settings can be found in Supplementary data. Subsequent to converting the raw data to the mzML format, an in-house program developed based on XCMS was employed for peak detection, extraction, alignment, and integration. Raw datasets were initially annotated to an in-house MS2 database with a cutoff of 0.3. Where no matches were found, raw datasets were annotated to MS1 in Metlin Metabolite Database. Adduct manner: [M+H]+ and [M+Na]+ were selected in positive mode (POS), [M-H]- and [M+FA-H]- in negative mode (NEG). Mass error value was less than 20 ppm.

2.3.2 Differentially accumulated metabolites (DAMs) enrichment analysis

For the signal intensity values of POS and NEG metabolites in each treatment group, orthogonal projections to latent structures discriminant analysis (OPLS-DA) was conducted to screen the differential metabolites between different groups. A permutation test was performed 200 times to assess the risk of over-fitting for the model (Fig. S5). DAMs were determined using variable importance in projection (VIP) > 1, with p value < 0.05, and FC > 2 used as a cutoff. These compounds identified in both modes were consolidated prior to KEGG pathway enrichment analysis.

2.4 Transcriptomic and metabolomic integration

Information on the DAMs and the corresponding unigenes was initially targeted using KEGG

compound and KEGG orthology Databases, where expression level of unigene with a $p < 0.05$ was compared with the changes in DAMs. The DAMs and their corresponding transcripts were mapped to KEGG pathway and visualized using “pathview” package in R (Luo and Brouwer, 2013).

2.5 Ovaries histology

Histological analysis of ovaries was conducted as described in Lai et al. (2019). The dissected ovaries were fixed in formalin solution (Sigma-Aldrich, USA) for 24h at 4°C. After fixation, dehydration of the ovaries was conducted in a series of ethanol solutions (70/80/95/100%). The processed ovaries were then embedded in paraffin. Serial sagittal sections (5 μm) were cut with a rotary microtome (Leica RM2016, Germany), mounted on the slides and allowed to dry for 24h at room temperature. Deparaffinization of the sectioned ovaries was performed in xylenes, followed by rehydration with a graded ethanol series (100/95/70%). The ovarian sections were stained with Mayer’s hematoxylin and eosin (H&E staining, Sigma-Aldrich, USA). Examination and imaging on the stained ovarian sections were performed at a magnification of 100× using a light microscope (Olympus BX53, Japan).

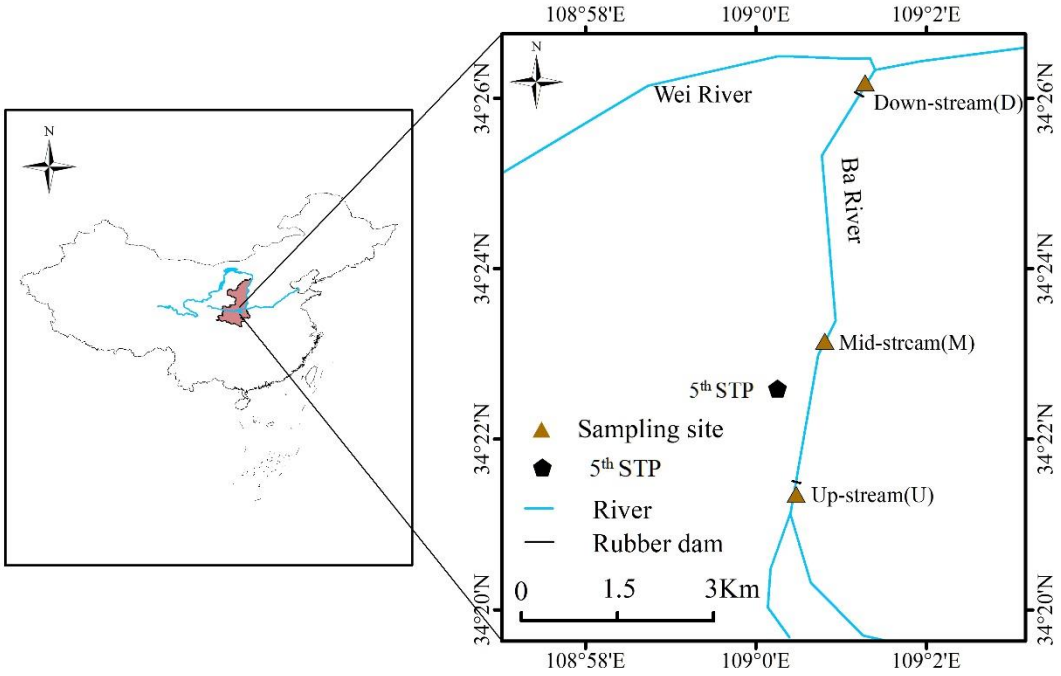

Fig. 1. Geographical distribution of three sampling sites in Ba River.

3 Results

3.1 Transcriptomic analysis

3.1.1 Illumina sequencing and *de novo* assembly

For sequencing data, the reads for the gonad samples ranged from 43.2 million to 51.9 million. After pruning the adapter and the reads with QV < 20, the rest clean data taking up approximately > 91% were assembled *de novo*, producing 146,384 unigenes with the mean length of 711.4 bp and a total length of 104.1 Mb.

3.1.2 Functional annotation

The unigenes functionally annotated to NR, GO, and KEGG databases occupied 31.16%, 9.63%, and 16.1%, respectively. Particularly, in the NR annotation, sharpbelly’s sequence exhibited high similarities to *Sinocyclocheilus rhinocerous* (18.98%), *Sinocyclocheilus*

anshuiensis (18.89%), *Sinocyclocheilus graham* (17.08%), and *Cyprinus carpio* (13.32%). Based on GO, 146,384 unigenes were enriched into 24 biological process (BP), 24 cellular components (CC), and 19 molecular function (MF) terms (Fig. S6). With respect to KEGG pathway, 35 pathways including 13 metabolism, 5 genetic information processing, 3 environmental information processing, 4 cellular processes, and 10 organismal systems were identified (Fig. S7).

3.1.3 Enrichment analyses of DEGs

A total of 238 (160 up- and 78 down-regulated), 255 (193 up- and 62 down-regulated), and 103 (38 up- and 65 down-regulated) genes were detected in U-M, D-M, and U-D groups, respectively. Three samples in the up- and two samples in the down-stream groups resembled that in the mid-stream, whereas others were clearly departed from the mid-stream group. As illustrated in heatmap, the gene expression patterns in sharpbelly collected from up- and down-stream groups were clearly distinguished from that in middle-stream. GO and pathway enrichment analyses were implemented for these DEGs to determine the terms that have been affected by the MWWE. The top 10 enriched GO terms concerning CC, BP, and MF were shown in Fig. S8. In addition to disruption on the human disease related processes, the DEGs were dominantly enriched in 9, 11, and 13 functional pathways in the U-M, D-M, and U-D groups, respectively. These pathways were associated with circulatory, digestive, endocrine, immune systems, environmental adaptation, lipid metabolism, signaling molecules interaction and transduction, and transport and catabolism.

Table 1: Comparison of fold changes detected for differentially accumulated metabolites (DAMs) and the unigenes with a *p* value less than 0.05.

KO	Gene	Log2 fold change	Compound ID	Metabolite	Log2 fold change
U-M					
K07817	*ptprn*	-2.41	C00009	phosphoric acid	0.42
K01915	*glul*	-1.41	C00009	phosphoric acid	0.42
K01081	*e3.1.3.5*	1.43	C00009	phosphoric acid	0.42
K17505	*ppm1k*	-4.2	C00009	phosphoric acid	0.42
K08745	*slc27a1_4*	1.2	C00009	phosphoric acid	0.42
K18993	*ubash3*	-2.02	C00009	phosphoric acid	0.42
K06927	*dph6*	-1.07	C00013	diphosphate	0.8
K00976	*fpgt*	2	C00013	diphosphate	0.8
K08745	*slc27a1_4*	1.2	C00013	diphosphate	0.8
K01897	*acsl*	1.29	C00013	diphosphate	0.8
K00252	*gcdh*	-1.37	C00016	flavin adenine dinucleotide	1.22
K01915	*glul*	-1.41	C00025	glutamate	-0.45
K00280	*loxl2_3_4*	1.37	C00078	tryptophan	0.93
K00994	*chpt1*	1.52	C00307	cytidine 5’-diphosphocholine	0.99
K00498	*cyp11a*	2.08	C01953	pregnenolone	0.85

D-M					
K00252	*gcdh*	-1.31	C00016	flavin adenine dinucleotide	0.65
K00252	*gcdh*	-1.31	C00016	flavin adenine dinucleotide	0.65
K00967	*pcyt2*	1.72	C00063	cytidine triphosphate	-2.4
K16860	*pld3_4*	-1.38	C00114	choline	0.39
U-D					
K12419	*fads6*	-2.15	C00006	beta-nicotinamide adenine dinucleotide phosphate	-0.47
K00844	*hk*	1.39	C00031	glucose	-0.81
K15449	*tyw1*	-2.42	C00073	methionine	0.61

3.2 Metabolomic analysis

In the gonad of female sharpbelly, the metabolomic analysis detected 4381 metabolites in positive mode and 5449 in negative mode. Specifically, in the positive mode 70, 55, and 22 DAMs were detected in the U-M, D-M, and U-D groups, respectively; in the negative mode 21, 19, and 10 DAMs were identified in the U-M, D-M, and U-D groups, respectively. The top 10 pathways enriched by the consolidated DAMs, primarily in relation to fundamental metabolic pathways, were illustrated in Fig. S9. However, only 12 metabolites mapped to KEGG compound database had the unigenes with a *p* value less than 0.05. Meanwhile, more than one gene has been targeted for certain metabolites (Table 1). After comparing the expression trends between transcripts and metabolites, an agreement was detected for only six compounds comprising diphosphate, glutamate, tryptophan, cytidine 5'-diphosphocholine, pregnenolone, and a-nicotinamide adenine dinucleotide phosphate (Table 1). Pregnenolone and its corresponding transcripts enclosed in the steroid hormone biosynthesis pathway were visualized in Supplementary data (Fig. S10).

3.3 Histopathological analysis

In ovaries of the majority of female sharpbelly collected at the upstream and downstream sites, chromatin nucleolar and perinucleolar oocytes were dominant, which indicates most female sharpbelly living upstream and downstream were at the stage 0 (Figs. 3A, C). In contrast, ovaries of most female sharpbelly collected at mid-stream were dominated by cortical alveoli oocytes, suggesting these ovaries were at stage 1 (Fig. 3B).

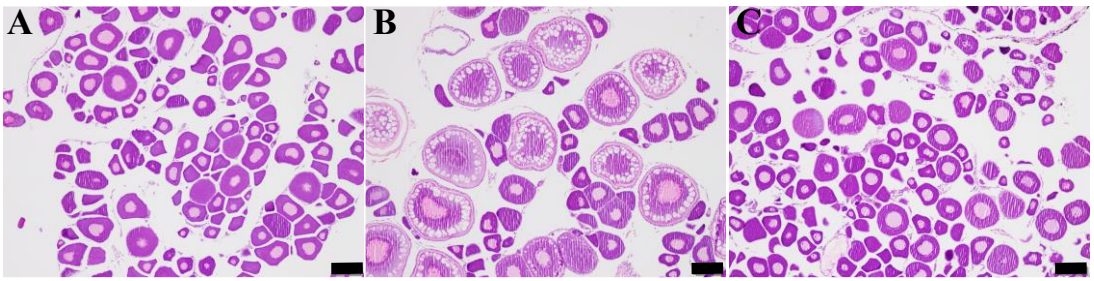

Fig. 3. Histological analysis of ovaries of female sharpbelly collected from the three sites in Ba River. (A) Ovaries of female sharpbelly collected in up-stream, (B) mid-stream, and (C) down-stream. A black bar represents 100 μm.

4. Publication

The paper will be submitted to Aquatic toxicology soon.

5. Perspectives in future

We will continue to analyze the effects of municipal wastewater effluents (MWWEs) on periphyton in Ba river.