


Draft genome of the Korean smelt *Hypomesus nipponensis* and its transcriptomic responses to heat stress in the liver and muscle

Biao Xuan,^{1,2} Jongbin Park,^{1,2} Sukjung Choi,² Inhwan You,^{1,2} Bo-Hye Nam,³ Eun Soo Noh,³ Eun Mi Kim,³ Mi-Young Song,⁴ Younhee Shin,⁵ Ji-Hyeon Jeon,^{5,6} and Eun Bae Kim ^{1,2,*}

¹Department of Applied Animal Science, College of Animal Life Sciences, Kangwon National University, Chuncheon 24341, Kangwon-do, Republic of Korea

²Laboratory of Microbial Genomics and Big Data, College of Animal Life Sciences, Kangwon National University, Chuncheon 24341, Kangwon-do, Republic of Korea

³Biotechnology Research Division, National Institute of Fisheries Science, Busan 46083, Republic of Korea

⁴Inland Fisheries Research Institute, National Institute of Fisheries Science, Gapyeong 12453, Republic of Korea

⁵Research and Development Center, Insilicogen Inc, Yongin 16954, Republic of Korea

⁶Department of Biological Science, Sungkyunkwan University, Suwon 16419, Republic of Korea

*Corresponding author: Department of Applied Animal Science, College of Animal Life Sciences, Kangwon National University, Chuncheon 200-701, Republic of Korea. Email: itanimal@kangwon.ac.kr

Abstract

Pond smelt (*Hypomesus nipponensis*) is a cold-freshwater fish species and a winter economic aquaculture resource in South Korea. Because of its high susceptibility to abnormal water temperature from global warming, a large number of smelt die in hot summers. Here, we present the first draft genome of *H. nipponensis* and transcriptomic changes in molecular mechanisms or intracellular responses under heat stress. We combined Illumina and PacBio sequencing technologies to generate the draft genome of *H. nipponensis*. Based on the reference genome, we conducted transcriptome analysis of liver and muscle tissues under normal (NT, 5°C) vs. warm (HT, 23°C) conditions to identify heat stress-induced genes and gene categories. We observed a total of 1987 contigs with N₅₀ of 0.46 Mbp, with the largest contig (3.03 Mbp) in the assembled genome. A total of 20,644 protein-coding genes were predicted, and 19,224 genes were functionally annotated: 15,955 genes for Gene Ontology terms and 11,560 genes for KEGG Orthology. We conducted the lost and gained genes analysis compared with three species that: human, zebrafish, and salmon. In the lost genes analysis, we detected that smelt lost 4461 (22.16%), 2825 (10.62%), and 1499 (3.09%) genes compare with above three species, respectively. In the gained genes analysis, we observed that smelt gained 1133 (5.49%), 1670 (8.09%), and 229 (1.11%) genes compared with the above species, respectively. From transcriptome analysis, a total of 297 and 331 differentially expressed genes (DEGs) with a false discovery rate <0.05 were identified in the liver and muscle tissues, respectively. Gene enrichment analysis of DEGs indicates that upregulated genes were significantly enriched for lipid biosynthetic process (GO:0008610, $P < 0.001$) and regulation of apoptotic process (GO:0042981, $P < 0.01$), and genes were downregulated by immune responses such as myeloid cell differentiation (GO:0030099, $P < 0.001$) in the liver under heat stress. In muscle tissue, upregulated genes were enriched for hypoxia (GO:0001666, $P < 0.05$), transcription regulator activity (GO:0140110, $P < 0.001$), and calcium-release channel activity (GO:0015278, $P < 0.01$), and genes were downregulated for a nicotinamide nucleotide biosynthetic process (GO:0019359, $P < 0.01$). The results of KEGG pathway analysis were similar to that of gene enrichment analysis. The draft genome and transcriptomic of *H. nipponensis* will be a useful genetic resource for functional and evolutionary studies. Our findings will improve understanding of molecular mechanisms and heat responses and be useful for predicting survival of the smelt and its closely related species under global warming.

Keywords: *Hypomesus nipponensis*; genome; transcriptome; high temperature; heat stress

Introduction

Pond smelt (*Hypomesus nipponensis*) is a member of the family Osmeridae and native to several countries, such as South Korea, China, Japan, and the United States. Smelt, a cold-freshwater fish species, is called “Bing eo” in Korea, which means ice-fish, because it resides in low-temperature water (0–15°C). It popular as a winter fishing species and as food in Korea, so many smelt festivals are held in winter. Smelt is an anadromous species (Saruwatari et al. 1997), but now it is cultivated in many reservoirs in South Korea. In a previous study, smelt samples from South Korea were identified as *H. nipponensis* (Choi and Kim 2019).

Abnormal increases in temperature lead to the death of many cold-water fish species. Therefore, many studies have focused on acute and chronic heat stress on fish (Logan and Buckley 2015). In Korea, there was local news report that the death of smelt in hot summers brings serious harm to the smelt industry. Therefore, understanding the characteristics of smelt will greatly improve responses to the problems faced. That said, however, there is no research on the genome or transcriptome of *H. nipponensis*. Furthermore, transcriptome analysis provides useful insight into the specific and general response to water temperature, such as heat shock response (Narum and Campbell

Received: March 23, 2021. Accepted: April 23, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

2015). In addition, the cellular response of different smelt species to high water temperature is different. In longfin smelt, many differentially expressed genes (DEGs) associated with heat shock proteins and chaperones are upregulated generally in fish responding to heat stress. Conversely, in Delta smelt, DEGs are associated with protein synthesis and metabolic processes (Basu et al. 2002). Cells generally regulate their own genes to survive under stress; however, occasionally they start programming cell death to induce apoptosis (Fulda et al. 2010). Increasing water temperature has many effects on fish, such as oxidative stress, endoplasmic reticulum (ER) stress, decreased immune function, hypoxia, and reproductive dysfunction that affects egg production and fertility (Olsvik et al. 2013; Lu et al. 2016; Qiang et al. 2017). Because specific tissue responses differ under heat stress, many heat stress-related studies used different tissues of fish, such as the liver, muscle, heart, kidney, brain, and gill (Logan and Buckley 2015). For example, ER stress was identified in Atlantic salmon liver under heat stress, where it promoted cell repair and reduced unfolded proteins; however, excessive ER stress led to cell apoptosis (Shi et al. 2019). ER stress also occurs in heat stress-affected rainbow trout (*Oncorhynchus mykiss*), where it is accompanied by changes in the immune system and post-transcriptional regulation of spliceosome (Huang et al. 2018). In teleost coho salmon (*Oncorhynchus kisutch*) liver, severe heat stressors can affect the redox state and induce oxidative stress (Nakano et al. 2014). Moreover, water temperature affects the concentration of dissolved oxygen in water, and low oxygen concentration in water is fatal to fish, leading to hypoxia. Dissolved oxygen plays an important role in maintaining biochemical and physiological processes. Hypoxia may limit energy processes and adversely affect growth, reproduction, and survival (Kelly et al. 2020). Many factors affect the growth and survival of fish; therefore, it is necessary to clarify the mechanism of heat stress responses of smelt to reduce their mortality effectively.

In the present study, we used sequencing technology to analyze the whole genome of *H. nipponensis* and transcriptome analysis to compare gene expression to understand the intracellular response mechanism of *H. nipponensis* at different temperatures in the liver and muscle tissues.

Materials and Methods

Ethics statement

All the procedures performed on animals were approved by the Institutional Animal Care and Use Committee (accept number: KW-181109-2) at Kangwon National University.

Source of *H. nipponensis* and genome sequencing

Whole-genome sequencing was conducted for three samples of smelt from Inje narincheon (South Korea, 38° 04' 34.0" N; 128° 11' 17.2" E), and the genomic DNA was extracted from the whole bodies using the DNeasy[®] Blood & Tissue kit (Qiagen, GmbH, Hilden, Germany). Illumina libraries for whole bodies were constructed using the TruSeq Nano Sample Prep kit (Illumina, San Diego, CA, USA) following the manufacturer's instructions. A PacBio Sequel platform library was constructed using the PacBio Template Prep kit. PacBio-generated reads were used for *de novo* assembly, and Illumina paired-end reads were mapped to the draft genome assembly to error correct by pilon v.1.23 (Walker et al. 2014). *De novo* assembly with PacBio reads was conducted by FALCON-UNZIP software version 1.2.5 using the parameters length_cutoff = 13 kb and length_cutoff_pr = 10 kb. *H. nipponensis* genome size was estimated using k-mer (17-, 19-, and 21-mers) analysis with Jellyfish 2.1.3 software. The assembly quality was assessed using Benchmarking Universal Single Copy Orthologs

(BUSCO) version 3.0 (Simão et al. 2015). Supplementary Figure S1 shows the analysis workflow of the fish genome.

PacBio (Iso-Seq) sequencing

Total RNA was isolated from liver and muscle tissues. For PacBio Iso-Seq data, we randomly selected 12 samples from the transcriptome analysis group (six liver and six muscle samples). The total RNA of the 12 samples was pooled for sequencing. The Iso-Seq library was prepared using a SMATer PCR cDNA Synthesis kit. We used Trinity (Haas et al. 2013) software to perform a genome-guided assembly and combine Iso-Seq data to predict the gene model.

H. nipponensis genome functional annotation

The predicted genes were subjected to search against the National Center for Biotechnology Information (NCBI) nonredundant protein database (20190306 ver.) using BLASTx, with an e-value cutoff of $1E^{-3}$. Gene name and description were assigned to each gene based on the highest hit with BLASTx. The genes were submitted to the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases to obtain the functional category and pathway information using Blast2GO and KAAS (Moriya et al. 2007), respectively.

Temperature trial and tissue sampling for RNA sequencing

Smelt were caught in the Hwacheon Dam (South Korea, 38° 07' 01.1" N; 127° 46' 43.1" E), and 100 fish were randomly placed in two 20-L tanks (50 fish per tank, NT and HT groups). The fish were fed, and acclimatization occurred for 2 h at a water (from Hwacheon Dam) temperature of 5°C. After acclimation, to simulate temperature conditions in their natural environment, one tank was heated from 5°C to 23°C at a constant rate of 4.5°C per h by hotrod, while the other tank was maintained at the same temperature (5°C) with ice packs. Oxygen was supplied throughout the experiment. Sampling began until loss of equilibrium among the fish population appeared in the HT group (23°C); however, no loss of equilibrium among the fish population was observed in the NT group (5°C). A total of 8 female fish were sampled (4 active and 4 inanimate from the NT and HT groups, respectively), and a total of 12 male fish were sampled (6 active and 6 inanimate from the NT and HT groups, respectively). After being sacrificed, muscle and liver tissues were immediately frozen in liquid nitrogen for gene expression profiling analysis.

Constructing mRNA libraries and sequencing

Tissue mRNA was extracted from the liver and muscle samples using TRIzol[®] reagent (Thermo Fisher Scientific, Waltham, MA, USA) and purified to remove DNA contamination using the TURBO DNA-free[™] kit (Invitrogen, Carlsbad, CA, USA). RNA concentration and quality were determined using an Epoch Microplate Spectrophotometer (BioTek, Winooski, VT, USA). The 37 sequencing libraries were created by reverse-transcription from 2 µg RNA from each sample using the TruSeq Stranded mRNA Library Prep kit (Illumina, San Diego, CA, USA). Index adapters were added to identify sequences for each sample in the final data. Subsequently, the 37 libraries were subjected to paired-end (2 × 101 bp) sequencing on the Illumina NovaSeq6000 Sequencing system.

Detection of lost and gained genes

To detect lost and gained genes in the Korean smelt genome, we compared their fragmented sequences (raw reads, assembled

DNA contigs, and assembled RNA transcripts) with several reference genomes: human (*Homo sapiens*), zebrafish (*Danio rerio*), and Atlantic salmon (*Salmo salar*). Sequence alignment was conducted by GASSST (v1.28) software. [Supplementary Figure S2](#) shows the method by which sequence alignment coverage (%) was calculated for each gene from the alignment. A low alignment coverage (<10%) indicates the absence of a gene in the genome.

Data analysis of mRNA

After removing the reads containing adaptor contamination using Trimmomatic v0.36, the clean reads were mapped to the assembled reference *H. nipponensis* genome using Bowtie v2.2.3. RSEM v1.2.31 was used to quantify gene abundances according to each *H. nipponensis* gene. Analysis of genes differentially expressed in the NT and HT groups was performed using the edgeR package ([Robinson et al. 2010](#)), and a false discovery rate (FDR) <0.05 was considered a threshold to identify significantly DEGs. GO enrichment analysis and KEGG statistical enrichment analysis of DEGs were performed using in-house Perl scripts (parent-child method) and the clusterProfiler R package ([Yu et al. 2012](#)), respectively. The significantly enriched GO terms were determined by Fisher's exact test, with a $P < 0.05$, and KEGG pathways were determined using FDR <0.05.

Results

Genome sequence data and genome size estimation

Using the PacBio long-read sequencing approach, a 126× coverage (58.60 Gbp) of *H. nipponensis* genome was obtained from >20-kb libraries. The short-read libraries (Illumina) generated an 80× coverage (36.99 Gbp) of genome, obtained from paired-end sequencing (2 × 101 bp) ([Supplementary Table S1](#)). The genome size was estimated to be 464 Mb using k-mer ($k = 17$) analysis ([Supplementary Table S2](#) and [Figure S3](#)).

Iso-Seq data generation

Isoform sequencing (PacBio) generated 0.49 Gb of data, which included 477,478 circular consensus sequences (CCS) reads and a mean 1023 CCS read length, yielding 18,799 transcripts ([@Supplementary Table S3](#)), which were used to generate the assistant file for *ab initio* gene model prediction.

De novo assembly of genome and quality assessment

Using long-read sequences from PacBio, the primary genome assembly was 631.85 Mbp in size, with 4106 contigs (largest contig of 3.03 Mbp, N_{50} of 0.32 Mbp, and L_{50} of 477). Evaluation of the genome completeness analysis against the 4584 genes of Actinopterygii BUSCO result showed that the 4290 genes (93.6%) were completely retrieved in the assembled genome, including 3911 single-copy genes (85.3%) and 379 duplicated genes (8.3%). In addition, 84 fragmented genes (1.8%) and 210 missing genes (4.5%) were found ([Supplementary Table S4](#)). After error correction using Pilon, the final improved primary genome assembly of 1987 contigs showed a total size of 499.59 Mbp, with an N_{50} of 0.46 Mbp, L_{50} of 300 contigs, GC content of 45.64%, and the largest contig length of 3.03 Mbp ([Supplementary Table S5](#)). [Supplementary Tables S6 and S7](#) show the Illumina and PacBio sequencing statistics. [Table 1](#) shows the *H. nipponensis* draft genome statistics.

Table 1 *Hypomesus nipponensis* draft genome statistics

Index	<i>H. nipponensis</i>
Number of contigs	1,987
Total contig length	498,930,205
Estimated genome size	464,167,205
Total length/estimated genome size	107.48%
Minimum contig length	19,619
Maximum contig length	3,031,274
Average contig length	251,097
Contig length N_{50}	464,523
GC contents	45.63%
BUSCO (Actinopterygii) complete	93.58%
BUSCO (Vertebrata) complete	95.32%
CEGMA complete	92.74%

Gene prediction and functional annotation

Evidence-based *ab initio* gene predictions with Augustus predicted 20,644 protein-coding genes. Among these, 93.12% (19,224) of predicted genes were annotated using the nonredundant database, while 6.88% (1420) of genes remained unannotated ([Supplementary Table S8](#)). [Figure 1A](#) shows the Blast top 10 hit species. Analysis of GO terms revealed that they were assigned to 15,955 genes on the three primary categories of ontology [biological process (BP), cellular component (CC), and molecular function (MF)]. [Figure 1B](#) shows the top 10 assigned GO terms for each category. A total of 11,560 genes were annotated with 7 categories on 447 pathways. [Figure 1C](#) shows the top three assigned KEGG terms for each category. [Table 2](#) shows the *H. nipponensis* consensus gene model.

Genome comparison for detection of lost and gained genes

We compared the genome of *H. nipponensis* with that of other species, including *H. sapiens*, *D. rerio*, and *S. salar*, for detection of lost and gained genes. For lost genes found from the alignment of the *H. nipponensis* genome compared with those of the other species, we compared 20,129, 26,612, and 48,548 genes in *H. sapiens*, *D. rerio*, and *S. salar*, respectively. We detected 4461 (22.16%), 2825 (10.62%), and 1499 (3.09%) genes with coverage less than 10% and 6047 (30.04%), 3901 (14.66%), and 2008 (4.14%) genes with coverage less than 15% ([Supplementary Table S9](#)). For gained genes found from the alignment of 20,644 genes of *H. nipponensis* compared with those of the other species, we observed 1133 (5.49%), 1670 (8.09%), and 229 (1.11%) genes with coverage less than 10% and 1911 (9.26%), 2484 (12.03%), and 362 (1.75%) genes with coverage less than 15% ([Supplementary Table S10](#)).

GO enrichment of lost and gained genes

Gene enrichment analysis comparing smelt with human showed that smelt lost some genes involved in stimulus, reproduction, nucleotide, and metal ions. Comparison with salmon showed that smelt lost chemotaxis-, immune system-, lipid-, mitochondria-, neuron-, nucleotide-, and reproduction-related genes ([Supplementary Table S11](#)). In gained genes enrichment analysis, compared with human, smelt gained chemotaxis-, heparin production-, immune system-, and stimulation-related genes. Compared with zebrafish and salmon, smelt gained chemotaxis- and immune system-related genes, respectively ([Supplementary Table S12](#)).

RNA-seq of liver and muscle transcriptome profiles at different temperature conditions

To identify differences in gene expression in smelt liver and muscle tissues at different temperatures, we constructed and sequenced 20

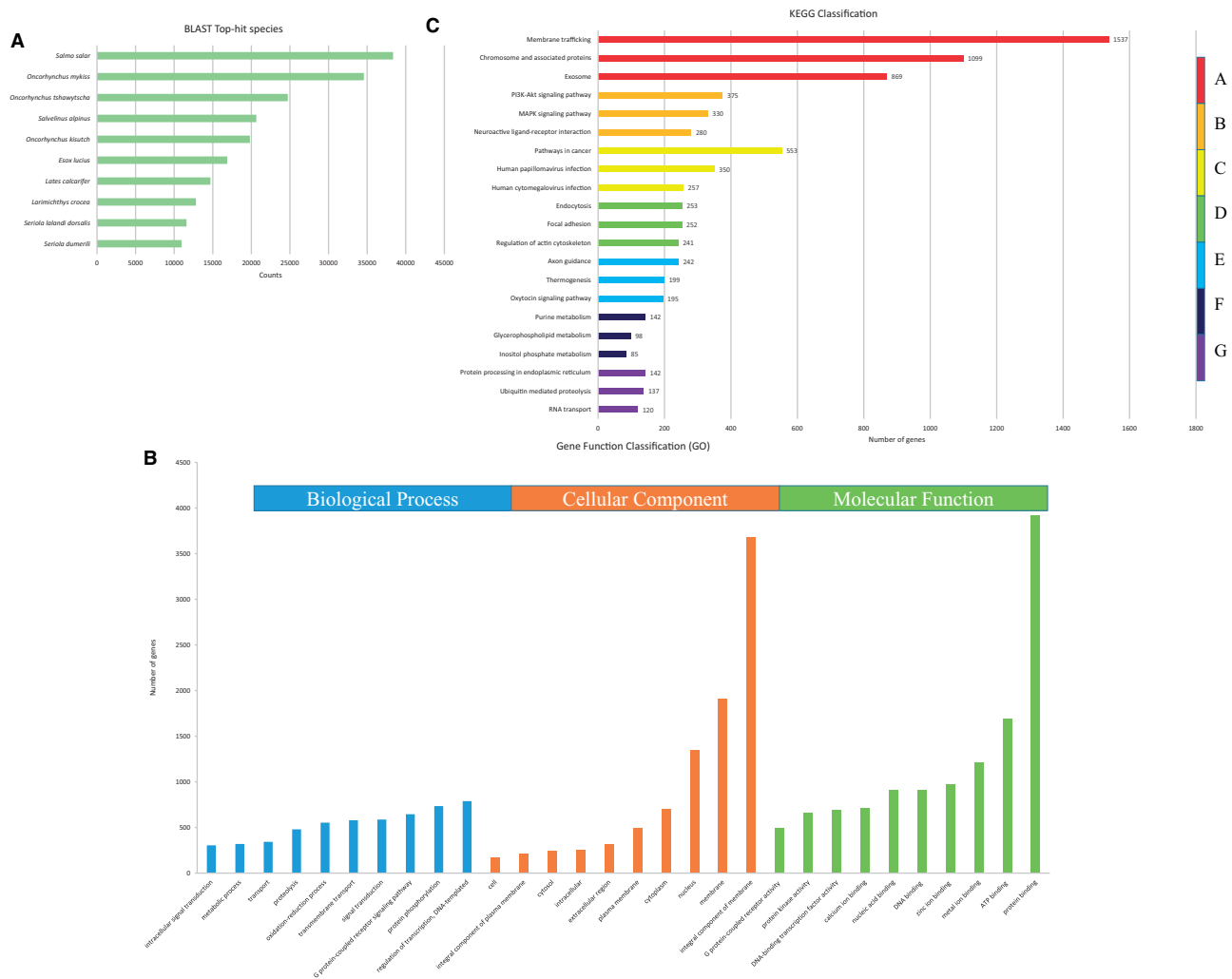


Figure 1 Annotation and functional classification of unigenes in the genome of *H. nipponensis*. (A) Blast top 10 hit species. (B) GO analysis was performed at level 2 for the three main categories (BP, CC, and MF). (C) Pathway assignment based on the KEGG database. Unigenes were classified into seven main categories (A: Genetic Information Processing; B: Metabolism; C: Organismal Systems; D: Cellular Processes; E: Human Diseases; F: Environmental Information Processing; G: Brite Hierarchies).

Table 2 *H. nipponensis* consensus gene model

Category	Index	Value
Gene	Gene count	20,644
	Maximum gene length	207,358 (HYNIP00072CG0010)
	Minimum gene length	222 (HYNIP00473CG0050)
	Average gene length	12,073.93
	Total gene length	249,254,225
	Genome coverage	49.96%
Exon	Exon count	193,124
	Exon count per gene	9.35
	Average exon length	178
	Total exon length	34,376,477
	Genome coverage	6.89%
	Intron	Intron count
Average intron length		1,245.81
Total intron length		214,877,748
Genome coverage		43.07%

cDNA libraries on the Illumina NovaSeq6000 platform with pair-end sequencing. The clean reads were generated by filtering the raw reads from the NT and HT groups. [Supplementary Tables S13 and S14](#) show the statistics of clean reads filtered from raw reads, and the mean quality is more than 30% for each sample.

Gene expression profiling

A total of 297 and 331 genes were significantly differentially expressed in liver and muscle tissues of female fish in the HT group compared with those in the NT group, respectively [$|\log_2(\text{Fold change})| > 1$, $\text{FDR} < 0.05$]. In the liver tissue of female fish, 297 DEGs, including 154 upregulated DEGs and 143 downregulated DEGs, were identified. In the muscle tissue of female fish, 331 DEGs, including 192 upregulated DEGs and 139 downregulated DEGs, were identified. In the liver tissue of male fish, eight DEGs, including four upregulated DEGs and four downregulated DEGs, were identified. In the muscle tissue of male fish, 29 DEGs, including 9 upregulated and 20 downregulated DEGs, were identified ([Figure 2](#)). The volcano plot showed the DEG analysis in the liver and muscle tissues of male fish. Because the number of DEGs in tissue from male fish is significantly less than that in tissue from female fish, the following gene functional enrichment analysis involves only tissue from female fish. [Supplementary Table S16](#) shows all DEGs.

GO analysis of DEGs in the liver and muscle

In enrichment analysis, DEGs were enriched GO terms (threshold as $P < 0.05$ and odds ratio > 1). For the liver, 128 upregulated and

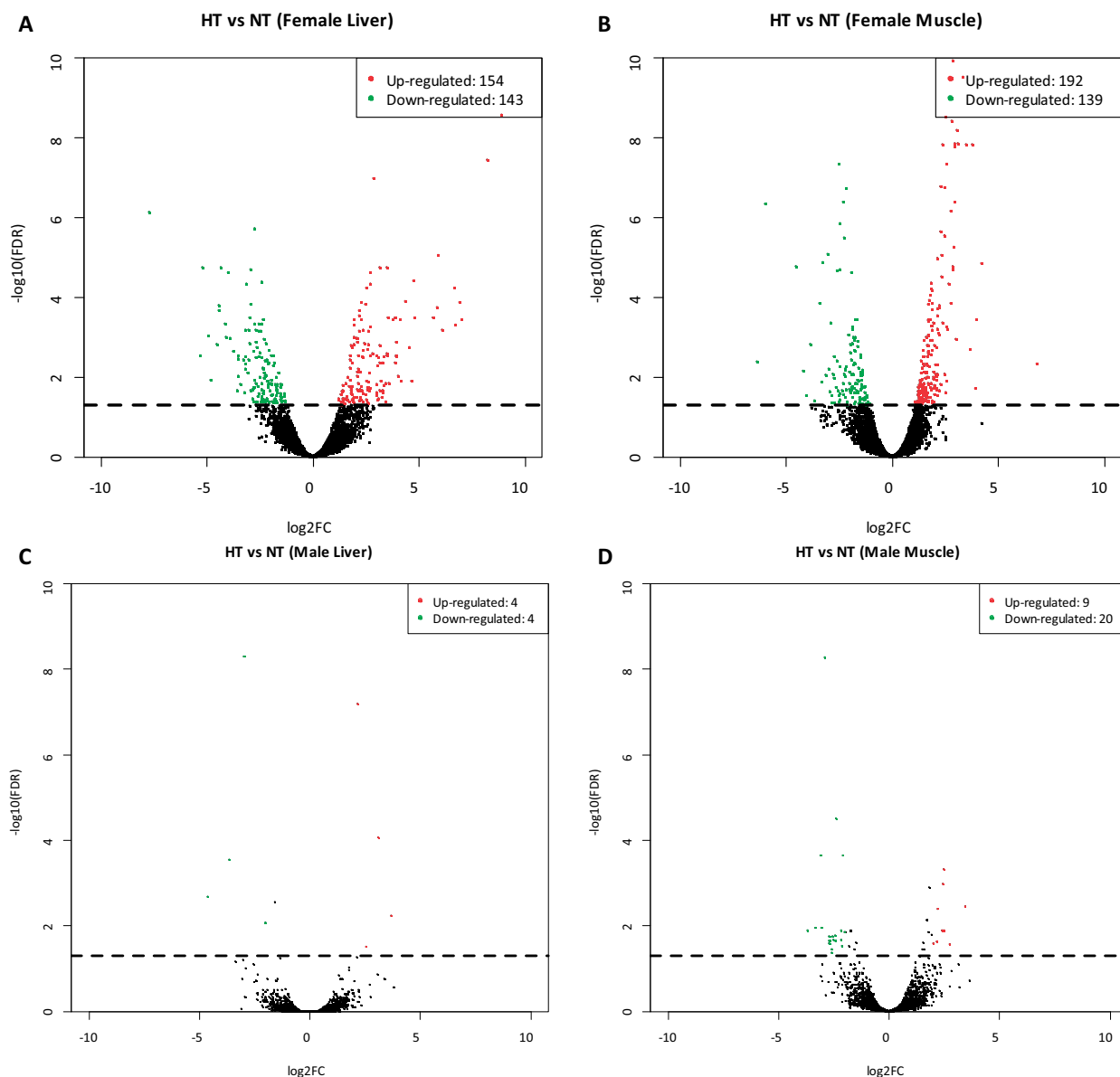


Figure 2 Differential expression of high temperature in *H. nipponensis* vs. low temperature. (A) DEGs in the liver of female fish. (B) DEGs in muscle of female fish. (C) DEGs in the liver of male fish. (D) DEGs in muscle of male fish. The dots above the dotted line are $FDR < 0.05$, $|\log_2FC| \geq 1$.

115 downregulated DEGs were annotated with GO terms. Upregulated GO terms included lipid metabolism-, cell death-, DNA binding-, and stress-related genes, and downregulated GO terms included immune system-, membrane protein-, and digestion-related genes (Table 3). For the muscle, 161 upregulated and 114 downregulated DEGs were annotated with GO terms. Upregulated GO terms included apoptosis-, hypoxia-, ion transportation-, transcription-, and component organization-related genes, and downregulated GO terms included nicotinamide nucleotide- and ubiquitination-related genes (Table 4).

KEGG pathway enrichment analysis of DEGs in the liver and muscle

Four KEGG pathways were enriched in the liver tissue. One pathway ($FDR < 0.1$), including the terpenoid backbone biosynthesis pathway, was upregulated, while three pathways ($FDR < 0.05$),

including the influenza A, Jak-STAT signaling, and pancreatic secretion pathways, were downregulated. Thirty-two KEGG pathways were enriched in the muscle tissue. Mitogen-activated protein kinase (MAPK) signaling ($FDR < 0.05$), oxytocin signaling pathways ($FDR < 0.05$), and HIF-1 signaling pathway ($FDR < 0.05$) were upregulated, while four pathways ($FDR < 0.05$), including inflammatory bowel disease (IBD), Leishmaniasis, African trypanosomiasis, and type 1 diabetes pathways, were downregulated (Figure 3).

Discussion

H. nipponensis is one of the most popular freshwater fish species in South Korean winters. With the rise in global warming, fish are suffering from heat stress, and a large number of deaths have been observed, resulting in great economic losses to the

Table 3 Gene enrichment analysis of DEGs in liver tissue under heat stress

	Category	GO ID	GO term	Bg_Gene	Tg_Gene	Bg_Gene	Tg_Gene	P-value	Odds ratio
Upregulated	Lipid metabolism	GO:0008610	Lipid biosynthetic process	15,955	128	127	7	0.00	6.87
		GO:0006629	Lipid metabolic process	15,955	128	371	10	0.00	3.36
		GO:0006720	Isoprenoid metabolic process	15,955	128	24	4	0.00	20.76
		GO:0008299	Isoprenoid biosynthetic process	15,955	128	13	3	0.00	28.74
		GO:0006721	Terpenoid metabolic process	15,955	128	16	2	0.01	15.57
		GO:0008202	Steroid metabolic process	15,955	128	35	4	0.00	14.24
	Cell death	GO:0016125	Sterol metabolic process	15,955	128	17	3	0.00	21.98
		GO:0006694	Steroid biosynthetic process	15,955	128	21	3	0.00	17.79
		GO:0042981	Regulation of apoptotic process	15,955	128	154	6	0.00	4.86
		GO:0043067	Regulation of programmed cell death	15,955	128	157	6	0.00	4.76
		GO:0010941	Regulation of cell death	15,955	128	164	6	0.00	4.56
		GO:0030983	Mismatched DNA binding	15,955	128	13	2	0.01	19.16
	DNA binding	GO:0006298	Mismatch repair	15,955	128	15	2	0.01	16.61
		GO:0003690	Double-stranded DNA binding	15,955	128	117	4	0.02	4.26
		GO:0080135	Regulation of cellular response to stress	15,955	128	49	3	0.01	7.63
Stress	GO:0080134	Regulation of response to stress	15,955	128	78	3	0.03	4.79	
	GO:0030099	Myeloid cell differentiation	15,955	115	44	4	0.00	12.60	
Downregulated	Immune system	GO:0030218	Erythrocyte differentiation	15,955	115	23	3	0.00	18.08
		GO:0048821	Erythrocyte development	15,955	115	10	2	0.00	27.69
		GO:0061515	Myeloid cell development	15,955	115	16	2	0.01	17.33
	Membrane protein	GO:0005834	Heterotrimeric G-protein complex	15,955	115	30	3	0.00	13.87
		GO:1905360	GTPase complex	15,955	115	30	3	0.00	13.87
	Digestion	GO:0007586	Digestion	15,955	115	13	2	0.01	21.33

Table 4 Gene enrichment analysis of DEGs in muscle tissue under heat stress

	Category	GO ID	GO term	Bg_Gene	Tg_Gene	Bg_Gene	Tg_Gene	P-value	Odds ratio
Upregulated	Apoptosis	GO:0097190	Apoptotic signaling pathway	15,955	161	16	3	0.00	18.56
		GO:0008630	Intrinsic apoptotic signaling pathway in response to DNA damage	15,955	161	8	2	0.00	24.74
	Hypoxia	GO:0097193	Intrinsic apoptotic signaling pathway	15,955	161	13	2	0.01	15.24
		GO:0036293	Response to decreased oxygen levels	15,955	161	17	2	0.02	11.65
	Ion transportation	GO:0001666	Response to hypoxia	15,955	161	17	2	0.02	11.65
		GO:0015278	Calcium-release channel activity	15,955	161	30	3	0.00	9.91
	Transcription	GO:0099604	Ligand-gated calcium channel activity	15,955	161	31	3	0.00	9.59
		GO:0019722	Calcium-mediated signaling	15,955	161	15	2	0.01	13.20
		GO:0003700	DNA-binding transcription factor activity	15,955	161	525	17	0.00	3.21
		GO:0140110	Transcription regulator activity	15,955	161	608	18	0.00	2.93
		GO::0000981	DNA-binding transcription factor activity, RNA polymerase II-specific	15,955	161	102	5	0.00	4.86
	Component organization	GO:0006355	Regulation of transcription, DNA-templated	15,955	161	1,038	21	0.00	2.00
		GO:0051639	Actin filament network formation	15,955	161	2	2	0.00	98.80
		GO:0032432	Actin filament bundle	15,955	161	4	2	0.00	49.42
		GO:0051017	Actin filament bundle assembly	15,955	161	22	3	0.00	13.50
GO:0061572		Actin filament bundle organization	15,955	161	24	3	0.00	12.38	
Downregulated	Nicotinamide nucleotide	GO:0019359	Nicotinamide nucleotide biosynthetic process	15,955	114	13	2	0.01	21.52
		GO:0046496	Nicotinamide nucleotide metabolic process	15,955	114	15	2	0.01	18.65
		GO:0019363	Pyridine nucleotide biosynthetic process	15,955	114	15	3	0.00	27.93
		GO:0019362	Pyridine nucleotide metabolic process	15,955	114	16	3	0.00	26.19
	Ubiquitination	GO:0072525	Pyridine-containing compound biosynthetic process	15,955	114	18	3	0.00	23.29
		GO:0072524	Pyridine-containing compound metabolic process	15,955	114	19	3	0.00	22.07
		GO:0061630	Ubiquitin protein ligase activity	15,955	114	73	3	0.02	5.75
GO:0061659	Ubiquitin-like protein ligase activity	15,955	114	73	3	0.02	5.75		

aquaculture industry. The present study reports the first high-quality draft genome assembly of *H. nipponensis* from the family Osmeridae (order Osmeriformes) and the analysis of transcriptome in heat stress to elucidate the mechanism of cellular response in liver and muscle tissues.

Genome sequencing

We successfully generated *H. nipponensis* draft genome assembly of 1987 highly contiguous contigs (498.3 Mbp) with a high N_{50} (0.47 Mbp). The estimated genome size of 464 Mbp was consistent

with that of another member of Osmeridae and greater than that of *Osmerus eperlanus* (European smelt, 342.8 Mbp) in the family Osmeridae (Malmström et al. 2016). Compared with other species (mammal, bird, bony fish, and cartilaginous fish), the genome size of smelt is smaller than that of other species, but the GC content of genome and coding sequence (CDS) and GC content of CDS are similar to those of other species, and the number of CDS is even more than that of other species (Supplementary Table S15). The genome size of *H. nipponensis* is within the range of most published fish genome sizes (Fan et al. 2020).

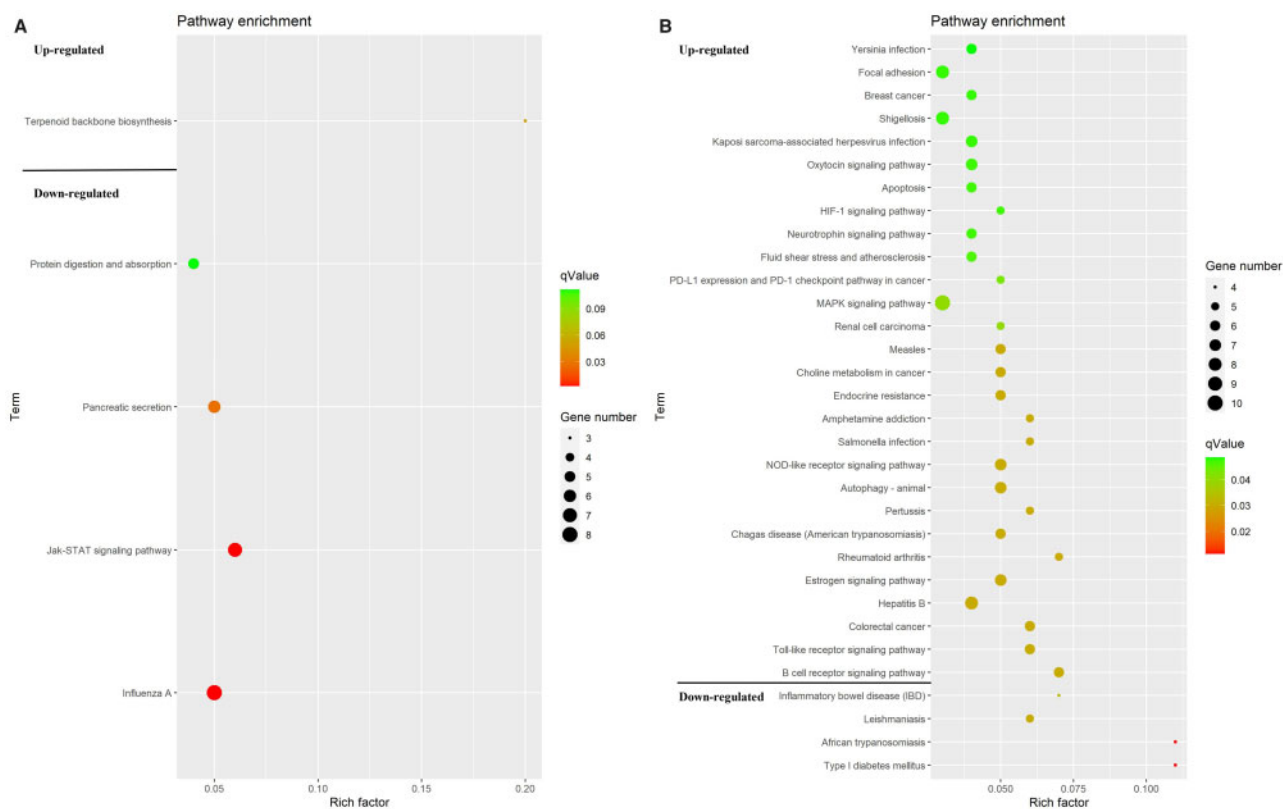


Figure 3 An overview of the KEGG pathway significantly enriched in DEGs in the (A) liver, and (B) muscle. The specific pathways are plotted along the y-axis, and the x-axis indicates the enrichment factor. The colored dots indicate the number of significant DEGs associated with each corresponding pathway: pathways with larger-sized dots contain a higher number of genes. The color of the each dot indicates the corrected P-value for the corresponding pathway.

Lost and gained genes

In the lost gene analysis, we compared *H. nipponensis* with human (*H. sapiens*), zebrafish (*D. rerio*), and Atlantic salmon (*S. salar*). The lowest number of lost genes was observed when comparing *H. nipponensis* with *S. salar* because salmon is a cold-water species that lives in similar water temperature. In addition, these two species are both anadromous; also, Osmeriformes are close relatives of the Salmoniformes. One of smelt species that rainbow smelt (*Osmerus mordax*) has high levels of similarity (86%) to salmonid genes (Von Schalburg et al. 2008). In the lost gene enrichment analysis, compared with human, smelt may not sense bitter taste because smelt eat phytoplankton; they may lose their taste buds for bitterness to survive. In one study comparing the naked mole rat (NMR) and human genomes, NMRs lost their receptors for bitter taste because NMRs live underground and their staple food is plant roots, most of which have a bitter taste, which facilitates this evolution (Kim et al. 2011). Moreover, the spermatogenesis of smelt may be different from that of human. With evolution, the genes involved in sex- and reproduction-related genes, such as mating behavior, fertilization, spermatogenesis, and sex determination, evolve with environmental changes (Volf 2005). The function of cytosine is less than that of human; ammonia may not be discharged smoothly when nitrogen is excreted. Compared with salmon, the sperm capacitation of smelt may be different from that of salmon. In gained genes functional enrichment analysis, compared with human, smelt have more genes related to heparin production, which may be because the living environment of smelt is low temperature, and heparin may prevent blood coagulation. Moreover, smelt may be more

sensitive to light than humans. Fish live in a different light environment from terrestrial species. However, water absorbs light, so as the water depth increases, the amount of light available decreases. Compared with zebrafish and salmon, the immune system of smelt is more complex.

Transcriptome analysis

The comparative gene expression profiles in the heat-stressed and normal groups are useful for understanding the mechanism of cellular responses under heat stress. In the HT group, when the temperature reached 23°C, smelt began to lose population equilibrium and die, and there were no dead fish in the NT group. The liver is a vital metabolic organ that relates to stress response. In the gene enrichment analysis, we found that some genes involved in lipid metabolism, such as the lipid biosynthetic, isoprenoid metabolic, and steroid metabolic processes. Lipid is the main component of cell membranes because acute heat stress may increase the fluidity of cell membranes, and the increase in lipid metabolism may be to maintain the stability of cell membranes. Also, lipid is the main component of fat, which can block external heat, and lipid metabolism may increase the synthesis of fat to block external heat. In the *Trematodus bernacchii* thermal acclimation study, there was an increase in membrane saturated fatty acids (Malekar et al. 2018). Moreover, lipids may help maintain energy homeostasis in fish because lipids are the basic components of sterols. Because of the membrane's unique molecular structure, it has a thermal-sensitive macromolecular structure. Environmental stress activates lipid metabolic enzymes and targets the downstream signaling pathway (Balogh et al. 2013). Poikilothermic organisms are able to maintain their membrane

fluidity for temperature fluctuation-induced cellular disturbance by regulating the composition of membrane lipids through physiologic and biochemical mechanisms of homeoviscous adaptation (Mendoza 2014). After heat shock, the plasma lipid peroxide level increases gradually, and severe heat stress affects the redox state and causes oxidative stress in salmon (Nakano et al. 2014). When fish suffer from hypoxia, fat metabolism is enhanced. Isoprenoid and terpenoid metabolic process-related genes, such as *HMGCS1*, *HMGCR*, and *FDPS*, were upregulated in liver tissue. These genes are involved in squalene synthesis. In a human cancer cell study, hypoxic cells displayed profound accumulation of squalene (Kucharzewska et al. 2015). Additionally, it plays an antioxidant role that eliminates ROS produced under stress (Micera et al. 2020). Squalene synthase is upregulated in the liver. It is the main material for the synthesis of sterols, and steroid metabolic process-related genes were upregulated in the liver under heat stress. In other studies, heat stress led to an increase in estrogen levels, and hormone disorders led to an imbalance in the number of males and females in the population (Shi et al. 2019). Some genes related to regulation of apoptotic process were significantly upregulated under heat stress. Although cells have different protective mechanisms under stress, the enhancement of stress can lead to cell signal interruption, extensive DNA damage, and cell apoptosis (Cheng et al. 2015). DNA damage is caused by hypoxia, leading to replication stress. Moreover, some genes related to regulation of cellular response to stress, such as *PLRG1*, *RTEL1*, and *ING2*, were upregulated. *PLRG1* is involved in pre-mRNA splicing as a component of the spliceosome critical for heat environment adaptation (Huang et al. 2018). *RTEL1* and *ING2* play a role in DNA repair under heat stress, indicating that liver tissue has replication problems under heat stress. However, some genes related to erythrocyte characteristics, including GO term erythrocyte differentiation, erythrocyte development, and regulation of anatomical structure size, were significantly downregulated under heat stress. Erythrocytes are cellular mediators of the immune response in teleost fish; thus, the immunity of smelt is reduced under heat stress. In addition, the genes related to cell membrane signal transduction were downregulated in the heat stress group, which indicates that it is difficult for cells to communicate with extracellular components under the condition of homeostasis imbalance. Moreover, digestion-related genes have been downregulated, suggesting that abnormal temperatures lead to reduced feeding, so fish should be fed less in hot summers. One study has shown that digestive enzyme activity was significantly affected by abnormal temperature (Pimentel et al. 2015). The effect of water temperature on digestive enzymes in fish depends on species, because the optimal temperature of enzyme activity is usually in the temperature range corresponding to fish habitat (Volkoff and Rønnestad 2020). Muscle is greatly affected by heat stress because in most species, this tissue constitutes approximately 50% of the body mass. Under heat stress, apoptosis-, hypoxia-, ion transportation-, transcription-, and component organization-related genes were upregulated. In high temperatures, muscle cells lose their calcium balance, which causes muscle spasm, and in muscle contraction-related genes in which those actin-filament genes were upregulated, hypoxia can also cause muscle contraction. In an oxygen consumption study, when the water temperature rises from 14°C to 20°C, the oxygen consumption of Delta smelt increased with the increase in temperature (Jeffries et al. 2016). In addition, nicotinamide adenine dinucleotide- and ubiquitination-related genes associated with cell respiration and energy metabolism were downregulated, possibly because of the decrease in collective energy metabolism caused by hypoxia.

In pathway enrichment analyses, under heat stress, the terpenoid backbone biosynthesis pathway in the liver was upregulated. The genes involved in this pathway are implicated in the final synthesis of squalene, which is an important enzyme under hypoxia conditions. Immune-related pathways, including the influenza A and Jak-STAT signaling pathways, were downregulated. Digestion-related pancreatic secretion, protein digestion, and absorption pathways were downregulated in the liver, which suggests that the digestive function of fish was disrupted under heat stress. In the muscle, several genes involved in the MAPK signaling pathway were upregulated under heat stress. A recent study has shown that the MAPK signaling pathway is activated in response to ER stress (Darling and Cook 2014). The disturbance of the ER environment, such as the decrease in Ca^{2+} concentration or the change in redox state, can affect protein folding and processing. Once misfolded proteins accumulate, ER stress activates a series of corresponding pathways. Under heat stress, some enriched genes in the oxytocin signaling pathway were also upregulated. Oxytocin activates the signal pathways of mRNA translation during ER stress (Klein et al. 2016). The HIF-1 signaling pathway was also upregulated in the muscle tissue, indicating that smelt suffered hypoxia stress in acute increasing temperatures. *HIF1A* is the main regulator of hypoxia-induced gene expression. The metabolism of smelt increases with the rise in temperature, which may lead to lack of oxygen. Smelt live in low-temperature environments, and the oxygen solubility in low-temperature water is higher than that in high-temperature water; thus, smelt is likely to have fewer red blood cells than temperate-water fish. Hypoxic stress-induced unfolded protein response (UPR) and autophagy. Hypoxia causes perturbations in ER activity, resulting in UPR activation. These upregulated pathways suggested that smelt suffered hypoxia and ER stress under heat stress. ER stress-related genes *NFE2L1* and *ERGIC1* were upregulated in the muscle tissue. ER stress was also observed under heat stress in Atlantic salmon (Shi et al. 2019). The IBD, Leishmaniasis, African trypanosomiasis, and type 1 diabetes pathways were downregulated in muscle under heat stress; these are related to immune and infectious diseases, considering that heat stress has a negative effect on the inner immune system (Lyu et al. 2018).

Conclusions

Using long reads and short reads from the PacBio Sequel and NovaSeq6000 platforms, respectively, we successfully assembled the draft genome and obtained the first reference genome of *H. nipponensis*. In transcriptome analysis, smelt suffer from hypoxia and ER stress, which leads to severe oxidative stress in the body under heat stress. These results provide a better understanding of the molecular mechanisms regulating the response of *H. nipponensis* under heat stress, which will help prevent and treat damage to fish caused by high water temperature.

Funding

This research was supported by a grant from the Collaborative Genome Program of the Korea Institute of Marine Science and Technology Promotion funded by the Ministry of Oceans and Fisheries (No. 20180430) and the National Institute of Fisheries Science (R2021038). Biao Xuan was supported by the BK21 Plus Program from Ministry of Education.

Data availability

H. nipponensis genome assembly was deposited in NCBI with accession number of SAMN16577099. RNA-Seq raw sequences were deposited in the Short Read Archive of the NCBI with accession number PRJNA672783. Supplementary material is available at figshare: <https://doi.org/10.25387/g3.14489730>.

Conflicts of interest

None declared.

Literature cited

- Balogh G, Péter M, Glatz A, Gombos I, Török Z, et al. 2013. Key role of lipids in heat stress management. *FEBS Lett.* 587:1970–1980. doi:10.1016/j.febslet.2013.05.016.
- Basu N, Basu N, Todgham AE, Todgham AE, Ackerman PA, et al. 2002. Heat shock protein genes and their functional significance in fish. *Gene.* 295:173–183.
- Cheng CH, Yang FF, Liao SA, Miao YT, Ye CX, et al. 2015. High temperature induces apoptosis and oxidative stress in pufferfish (*Takifugu obscurus*) blood cells. *J Therm Biol.* 53:172–179. doi:10.1016/j.jtherbio.2015.08.002.
- Choi S, Kim EB. 2019. Complete mitochondrial genome sequence and SNPs of the Korean smelt *Hypomesus nipponensis* (Osmeriformes, Osmeridae). *Mitochondrial DNA Part B Res.* 4:1844–1845. doi:10.1080/23802359.2019.1613178.
- Darling NJ, Cook SJ. 2014. The role of MAPK signalling pathways in the response to endoplasmic reticulum stress. *Biochim Biophys Acta.* 1843:2150–2163. doi:10.1016/j.bbamcr.2014.01.009.
- Fan G, Song Y, Yang L, Huang X, Zhang S, et al. 2020. Initial data release and announcement of the 10,000 Fish Genomes Project (Fish10K). *GigaScience.* 9. doi:10.1093/gigascience/giaa080.
- Fulda S, Gorman AM, Hori O, Samali A. 2010. Cellular stress responses: cell survival and cell death. *Int J Cell Biol.* 2010:214074. doi:10.1155/2010/214074.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8:1494–1512. doi:10.1038/nprot.2013.084.
- Huang J, Li Y, Liu Z, Kang Y, Wang J. 2018. Transcriptomic responses to heat stress in rainbow trout *Oncorhynchus mykiss* head kidney. *Fish Shellfish Immunol.* 82:32–40. doi:10.1016/j.fsi.2018.08.002.
- Jeffries KM, Connon RE, Davis BE, Komoroske LM, Britton MT, et al. 2016. Effects of high temperatures on threatened estuarine fishes during periods of extreme drought. *J Exp Biol.* 219:1705–1716. doi:10.1242/jeb.134528.
- Kelly T, Johnsen H, Burgerhout E, Tveiten H, Thesslund T, et al. 2020. Low oxygen stress during early development influences regulation of hypoxia-response genes in farmed Atlantic salmon (*Salmo salar*). G3 (Bethesda). 10:3179–3188. doi:10.1534/g3.120.401459.
- Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, et al. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature.* 479:223–227. doi:10.1038/nature10533.
- Klein BY, Tamir H, Hirschberg DL, Ludwig RJ, Glickstein SB, et al. 2016. Oxytocin opposes effects of bacterial endotoxin on ER-stress signaling in Caco2BB gut cells. *Biochim Biophys Acta.* 1860:402–411. doi:10.1016/j.bbagen.2015.10.025.
- Kucharzewska P, Christianson HC, Belting M. 2015. Global profiling of metabolic adaptation to hypoxic stress in human glioblastoma cells. *PLoS One.* 10:e0116740. doi:10.1371/journal.pone.0116740.
- Logan CA, Buckley BA. 2015. Transcriptomic responses to environmental temperature in eurythermal and stenothermal fishes. *J Exp Biol.* 218:1915–1924. doi:10.1242/jeb.114397.
- Lu Y, Wu Z, Song Z, Xiao P, Liu Y, et al. 2016. Insight into the heat resistance of fish via blood: effects of heat stress on metabolism, oxidative stress and antioxidant response of olive flounder *Paralichthys olivaceus* and turbot *Scophthalmus maximus*. *Fish Shellfish Immunol.* 58:125–135. doi:10.1016/j.fsi.2016.09.008.
- Lyu L, Wen H, Li Y, Li J, Zhao J, et al. 2018. Deep transcriptomic analysis of black rockfish (*Sebastes schlegelii*) provides new insights on responses to acute temperature stress. *Sci Rep.* 8. doi:10.1038/s41598-018-27013-z.
- Malekar VC, Morton JD, Hider RN, Cruickshank RH, Hodge S, et al. 2018. Effect of elevated temperature on membrane lipid saturation in Antarctic notothenioid fish. *PeerJ.* 6:e4765. doi:10.7717/peerj.4765.
- Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, et al. 2016. Evolution of the immune system influences speciation rates in teleost fishes. *Nat Genet.* 48:1204–1210. doi:10.1038/ng.3645.
- Mendoza DD. 2014. Temperature sensing by membranes. *Annu Rev Microbiol.* 68:101–116. doi:10.1146/annurev-micro-091313-103612.
- Micera M, Botto A, Geddo F, Antoniotti S, Bertera CM, et al. 2020. Squalene: more than a step toward sterols. *Antioxidants.* 9:688–614. doi:10.3390/antiox9080688.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35:W182–W185. doi:10.1093/nar/gkm321.
- Nakano T, Kameda M, Shoji Y, Hayashi S, Yamaguchi T, et al. 2014. Effect of severe environmental thermal stress on redox state in salmon. *Redox Biol.* 2:772–776. doi:10.1016/j.redox.2014.05.007.
- Narum SR, Campbell NR. 2015. Transcriptomic response to heat stress among ecologically divergent populations of redband trout. *BMC Genomics.* 16. doi:10.1186/s12864-015-1246-5.
- Olsvik PA, Vikeså V, Lie KK, Hevrøy EM. 2013. Transcriptional responses to temperature and low oxygen stress in Atlantic salmon studied with next-generation sequencing technology. *BMC Genomics.* 14:817. doi:10.1186/1471-2164-14-817.
- Pimentel MS, Faleiro F, Diniz M, Machado J, Pousão-Ferreira P, et al. 2015. Oxidative stress and digestive enzyme activity of flatfish larvae in a changing ocean. *PLoS One.* 10:e0134082. doi:10.1371/journal.pone.0134082.
- Qiang J, Bao WJ, Tao FY, He J, Li XH, et al. 2017. The expression profiles of miRNA-mRNA of early response in genetically improved farmed tilapia (*Oreochromis niloticus*) liver by acute heat stress. *Sci Rep.* 7:1–15. doi:10.1038/s41598-017-09264-4.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26:139–140. doi:10.1093/bioinformatics/btp616.
- Saruwatari T, Lopez JA, Pietsch TW. 1997. A revision of the Osmerid genus *Hypomesus* Gill (Teleostei: Salmoniformes), with the description of a new species from the Southern Kuril Islands. *SpecDiv.* 2:59–82. doi:10.12782/specdiv.2.59.
- Shi KP, Dong SL, Zhou YG, Li Y, Gao QF, et al. 2019. RNA-seq reveals temporal differences in the transcriptome response to acute heat stress in the Atlantic salmon (*Salmo salar*). *Comp Biochem Physiol Part D Genomics Proteomics.* 30:169–178. doi:10.1016/j.cbd.2018.12.011.

- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31: 3210–3212. doi:10.1093/bioinformatics/btv351.
- Volff JN. 2005. Genome evolution and biodiversity in teleost fish. *Heredity (Edinb)*. 94:280–294. doi:10.1038/sj.hdy.6800635.
- Volkoff H, Rønnestad I. 2020. Effects of temperature on feeding and digestive processes in fish. *Temperature (Austin)*. 7:307–320. doi:10.1080/23328940.2020.1765950.
- Von Schalburg KR, Leong J, Cooper GA, Robb A, Beetz-Sargent MR, et al. 2008. Rainbow smelt (*Osmerus mordax*) genomic library and EST resources. *Mar Biotechnol (NY)*. 10:487–491. doi:10.1007/s10126-008-9089-6.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 9:e112963. doi:10.1371/journal.pone.0112963.
- Yu G, Wang LG, Han Y, He QY. 2012. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 16: 284–287. doi:10.1089/omi.2011.0118.

Communicating editor: B. Andrews