bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1

2

3 # Draft Genome of the Korean smelt *Hypomesus nipponensis* and its transcriptomic
4 # responses to heat stress in the liver and muscle

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

8 [1] Department of Applied Animal Science, College of Animal Life Sciences, Kangwon National
9 University, Chuncheon 24341, Kangwon-do, Republic of Korea

10 [2] Laboratory of Microbial Genomics and Big Data, College of Animal Life Sciences, Kangwon
11 National University, Chuncheon 24341, Kangwon-do, Republic of Korea

12 [3] Biotechnology Research Division, National Institute of Fisheries Science, Busan 46083, Korea

13 [4] Inland Fisheries Research Institute, National Institute of Fisheries Science, Gapyeong 12453,
14 Korea

15 [5] Research and Development Center, Insilicogen Inc, Yongin 16954, Republic of Korea

16 [6] Department of Biological Science, Sungkyunkwan University, Suwon 16419, Korea

17

18 **# Corresponding author**

19 **Mailing address:** Department of Applied Animal Science, College of Animal Life Sciences,
20 Kangwon National University, Chuncheon 200-701, Republic of Korea.

21 **Tel:** +82-33-250-8642

22 **Fax:** +82-33-259-5574

23 **E-mail:** itanimal@kangwon.ac.kr

24

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Pond smelt (*Hypomesus nipponensis*) is a cold-freshwater fish species as a winter economic resource of aquaculture in South Korea. Due to its high susceptibility to abnormal water temperature from global warming, a large number of smelt die in hot summer. 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℃) versus warm (HT, 23℃) conditions, to identify heat stress-induced genes and gene categories. We observed a total of 1,987 contigs, with N_{50} of 0.46 Mbp with a largest contig (3.03 Mbp) in the assembled genome. A total number of 20,644 protein coding genes were predicted, and 19,224 genes were functionally annotated: 15,955 genes for Gene Ontology (GO) terms; and 11,560 genes for KEGG Orthology (KO). We conducted the lost and gained genes analysis compared with three species that human, zebrafish and salmon. In the lost genes analysis, we detected smelt lost 4,461 (22.16%), 2,825 (10.62%), and 1,499 (3.09%) genes compare with above three species, respectively. In the gained genes analysis, we observed smelt gain 1,133 (5.49%), 1,670 (8.09%), and 229 (1.11%) genes compare with above species, respectively. From transcriptome analysis, a total of 297 and 331 differentially expressed genes (DEGs) with False discovery rate (FDR) < 0.05 were identified in the liver and muscle tissues, respectively. Gene enrichment analysis of DEGs indicates that up-regulated genes were significantly enriched for lipid biosynthetic process (GO:0008610, $P < 0.001$) and regulation of apoptotic process (GO:0042981, $P < 0.01$), and down-

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

regulated genes by immune responses such as myeloid cell differentiation (GO:0030099, $P < 0.001$) in the liver under heat stress. In muscle tissue, up-regulated 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 down-regulated genes for 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 used as a useful genetic resource for functional and evolutionary studies. Our findings will improve understanding of the molecular mechanisms and heat responses and will 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

69 **Introduction**

70 Pond smelt (*Hypomesus nipponensis*) is a member of the family Osmeridae native to
71 several countries, such as South Korea, China, Japan and the United States. Smelt as a
72 cold-freshwater fish species it is called "Bing eo" in Korea, which means ice-fish,
73 because it resides in low-temperature water that 0 to 15°C, and it is very popular as
74 winter fishing and food in Korea, so there are many smelt festivals are held in winter.
75 Smelt is an anadromous species(Saruwatari, Lopez, & Pietsch, 1997), however, now it is
76 cultivated in many reservoirs in South Korea. In a previous study, smelt samples from
77 South Korea were identified as *H. nipponensis* (Choi & Kim, 2019).

78 The abnormal increase of temperature leads to the death of a large number of cold-water
79 fish species. Therefore, many studies have focused on the acute and chronic heat stress of
80 fish(Logan & Buckley, 2015). In Korea, there are local news report that the death of
81 smelt in hot summer brings serious harm to the smelt industry. Therefore, understanding
82 the characteristics of smelt will greatly improve the problems faced, however, there is no
83 research on the genome or transcriptome of *H. nipponensis*. Furthermore, transcriptome
84 analysis provides useful insight into the specific and general response to water
85 temperature, such as heat shock response(Narum & Campbell, 2015). In addition, the
86 cellular response of different smelt species to high water temperature is different. In
87 longfin smelt, many DEGs associated with heat shock proteins and chaperones that up-
88 regulated generally in fish response to heat stress. Conversely, in Delta smelt, DEGs
89 associated with protein synthesis and metabolic processes(Basu et al., 2002). Cells
90 generally regulate their own genes to survive under stress; however, occasionally they
91 start programming cell death to induce apoptosis(Fulda, Gorman, Hori, & Samali, 2010).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

92 Increasing water temperature has many effects on fish such as oxidative stress,
93 endoplasmic reticulum (ER) stress, decreased immune function, hypoxia, and
94 reproductive dysfunction affecting egg production and fertility(Lu et al., 2016; Olsvik,
95 Vikeså, Lie, & Hevrøy, 2013; Qiang et al., 2017). Since specific responses of different
96 tissues were different under heat stress, many heat stress-related studies used different
97 tissues of fish such as the liver, muscle, heart, kidney, brain, and gill (Logan & Buckley,
98 2015). For example, ER stress was identified in Atlantic salmon liver under heat stress,
99 while ER stress promoted cell repair and reduced unfolded proteins; however, excessive
100 ER stress led to cell apoptosis (Shi et al., 2019). ER stress also occurs in heat stress-
101 treated rainbow trout *Oncorhynchus mykiss*, and it is accompanied by changes in the
102 immune system and post-transcriptional regulation of spliceosome (Huang, Li, Liu, Kang,
103 & Wang, 2018). In teleost coho salmon (*Oncorhynchus kisutch*) liver, severe heat
104 stressors can affect the redox state and induce oxidative stress (Nakano et al., 2014).
105 Moreover, water temperature affects the concentration of dissolved oxygen in water, low-
106 oxygen concentration in water is fatal to fish, leading to hypoxia, dissolved oxygen plays
107 an important role in maintaining biochemical and physiological processes. Hypoxia may
108 limit energy processes and adversely affect growth, reproduction and survival(Kelly et al.,
109 2020). There are many factors affecting the growth and survival of fish, therefore, it is
110 necessary to clarify the mechanism of heat stress responses of smelt to reduce the
111 mortality effectively.

112 In the present study, we used sequencing technology to analyze the whole genome of *H.*
113 *nipponensis* and transcriptome analysis to compare gene expression to understand the

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

114 intracellular response mechanism of *H. nipponensis* at different temperatures in the liver
115 and muscle tissues.

116 Materials and Methods

117 Ethics statement

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

121 Source of *H. nipponensis* and genome sequencing

122 Whole genome sequencing was conducted for 3 samples of smelt from Inje narincheon
123 (South Korea, 38° 04' 34.0" N; 128° 11' 17.2" E) and the genomic DNA was extracted
124 from the whole bodies using DNeasy® Blood & Tissue kit (Qiagen, GmbH, Hilden,
125 Germany). Illumina libraries for whole bodies were constructed using the TruSeq Nano
126 Sample Prep kit (Illumina, San Diego, CA, USA) following the manufacturer's
127 instructions. PacBio Sequel platform library was constructed using the PacBio Template
128 Prep kit. PacBio generated reads were used for *de novo* assembly and Illumina paired-end
129 reads were mapped to the draft genome assembly to error correct by pilon v.1.23(Walker
130 et al., 2014). *de novo* assembly with PacBio reads was conducted by FALCON-UNZIP
131 software 1.2.5 version, using the parameters length_cutoff = 13kb and length_cutoff_pr =
132 10kb. *H. nipponensis* genome size was estimated using k-mer (17-, 19- and 21-mers)
133 analysis with Jellyfish 2.1.3 software. The assembly quality was assessed using
134 Benchmarking Universal Single Copy Orthologs (BUSCO) v.3.0(Simão, Waterhouse,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

135 Ioannidis, Kriventseva, & Zdobnov, 2015). Figure S1 shows the analysis workflow of
136 fish genome.

137 **PacBio (Iso-Seq) sequencing**

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

144 ***H. nipponensis*** **genome functional annotation**

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

152 **Temperature trial and tissue sampling for RNA sequencing**

153 Smelt were caught in the Hwacheon Dam (South Korea, 38° 07' 01.1" N; 127° 46' 43.1"
154 E) and one hundred fish were randomly placed in two 20-L tanks (50 fish per tank, NT
155 and HT groups). The fish were fed, and acclimatization occurred for 2 h at a water (from
156 Hwacheon Dam) temperature of 5°C. After acclimation, to simulate temperature

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

157 condition in natural environment, one tank was heated from 5°C to 23°C at a constant
158 rate of 4.5°C per h by hotrod, while the other tank was maintained at the same
159 temperature (5°C) with ice pack. Oxygen was supplied throughout the experiment.
160 Sampling began until loss of equilibrium fish population appeared in the HT group
161 (23°C); however, no loss of equilibrium fish population was observed in the NT group
162 (5°C). A total of eight female fish were sampled (four active and four inanimate from NT
163 and HT groups, respectively) and a total of twelve male fish were sampled (six active and
164 six inanimate from NT and HT groups, respectively). After being sacrificed, muscle and
165 liver tissues were immediately frozen in liquid nitrogen for gene expression profiling
166 analysis.

167 **Constructing mRNA libraries and sequencing**

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

178 **Detection of lost and gained genes**

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

179 To detect lost and gained genes in the Korean smelt genome, we compared their
180 fragmented sequences (raw reads, assembled DNA contigs and assembled RNA
181 transcripts) with several reference genome of that the human (*Homo sapiens*), zebrafish
182 (*Danio rerio*) and Atlantic salmon (*Salmo salar*). Sequence alignment was conducted by
183 GASSST (v1.28) software. Figure S2 shows the method that sequence alignment
184 coverage (%) was calculated for each gene from the alignment. As a low alignment
185 coverage (less than 10%) indicates absence of a gene in the genome.

186 **Data analysis of mRNA**

187 After removing the reads containing adaptor contamination using Trimmomatic v0.36,
188 the clean reads were mapped to the assembled reference *H. nipponensis* genome using
189 Bowtie v2.2.3. RSEM v1.2.31 was used to quantify gene abundances according to each *H.*
190 *nipponensis* gene. Analysis of genes differentially expressed in the NT and HT groups
191 was performed using the edgeR package (Robinson, McCarthy, & Smyth, 2009), and
192 FDR < 0.05 was considered as a threshold to identify significantly differentially
193 expressed genes (DEGs). GO enrichment analysis and KEGG statistical enrichment
194 analysis of DEGs were performed using in-house perl scripts (parent-child method) and
195 clusterProfiler R package (Yu, Wang, Han, & He, 2012), respectively. The significantly
196 enriched GO terms were determined by Fisher's exact test with a $p < 0.05$, and KEGG
197 pathways were determined using FDR < 0.05.

198 **Data accessibility**

199 *Hypomesus nipponensis* genome assembly was deposited in National Center for
200 Biotechnology Information (NCBI) with accession number of SAMN16577099. RNA-

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

201 Seq raw sequences were deposited in the Short Read Archive of the NCBI with accession
202 number PRJNA672783.

203

204 ## Results

205 ### Genome sequence data and genome size estimation

206 Using the PacBio long-read sequencing approach, a 126× coverage (58.60 Gbp) of *H.*
207 *nipponensis* genome was obtained from > 20-kb libraries. The short-read libraries
208 (Illumina) generated an 80× coverage (36.99 Gbp) of genome was obtained from paired
209 end sequencing (2x101bp) (Table S1). The genome size was estimated to be 464 Mb
210 using k-mer (k=17) analysis (Table S2, Figure S3).

211 ### Iso-Seq data generation

212 Isoform sequencing (PacBio) generated 0.49 Gb of data, which included 477,478 circular
213 consensus sequences (CCS) reads and 1,023 mean of CCS read length, yielding 18,799
214 transcripts (Table S3), which were used to generate the assistant file for *ab initio* gene
215 model prediction.

216 ### *De novo* assembly of genome and quality assessment

217 Using long-read sequences from PacBio, the primary genome assembly was 631.85 Mbp
218 in size, with 4,106 contigs (largest contig of 3.03 Mbp, N_{50} of 0.32 Mbp, and L_{50} of 477).
219 Evaluation of the genome completeness analysis, against the 4,584 genes of
220 *Actinopterygii*, BUSCO result shows the 4,290 genes (93.6%) were completely retrieved
221 in the assembled genome, including 3,911 single-copy genes (85.3%) and 379 duplicated

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

222 genes (8.3%). In addition, 84 fragmented genes (1.8%) and 210 missing genes (4.5%)
223 were found (Table S4). After error correction using pilon, the final improved primary
224 genome assembly of 1,987 contigs showed a total size of 499.59 Mbp, with an N_{50} of
225 0.46 Mbp, L_{50} of 300 contigs, GC content of 45.64%, and the largest contig length of
226 3.03 Mbp (Table S5). Table S6 and S7 show the Illumina and PacBio sequencing
227 statistics. Table 1 shows the *H. nipponensis* draft genome statistics.

228 **Gene prediction and functional annotation**

229 Evidence-based *ab initio* gene predictions with Augustus predicted 20,644 protein-coding
230 genes. Among these, 93.12% (19,224) predicted genes were annotated using the non-
231 redundant database, while 6.88% (1,420) genes remained unannotated (Table S8). Figure
232 1.A shows the Blast top 10 hit species. Analysis of GO terms revealed that they were
233 assigned to 15,955 genes on the three primary categories of ontology (biological process
234 (BP), cellular component (CC), and molecular function (MF)). Figure 1B shows the top
235 10 assigned GO terms for each category. A total of 11,560 genes were annotated with
236 seven categories on 447 pathways. Figure 1C shows the top three assigned KEGG terms
237 for each category. Table 2 shows the *H. nipponensis* consensus gene model.

238 **Genome comparison for detection of lost and gained genes**

239 We compared the genome of *H. nipponensis* with that of other species, including *Homo*
240 *sapiens*, *Danio rerio*, and *Salmo salar* for detection of lost and gained genes. For lost
241 genes found from the alignment of *H. nipponensis* genome to those of the other species,
242 we compared 20,129, 26,612, and 48,548 genes in *H. sapiens*, *D. rerio*, and *S. salar*,
243 respectively. We detected 4,461 (22.16%), 2,825 (10.62%), and 1,499 (3.09%) genes

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

244 with coverage less than 10% and 6,047 (30.04%), 3,901 (14.66%), and 2,008 (4.14%)
245 genes with coverage less than 15% (Table S9). For gained genes found from the
246 alignment of 20,644 genes of *H. nipponensis* to those of the other species, we observed
247 1,133 (5.49%), 1,670 (8.09%), and 229 (1.11%) genes with coverage less than 10% and
248 1,911 (9.26%), 2,484 (12.03%), and 362 (1.75%) genes with coverage less than 15%
249 (Table S10).

250 **GO enrichment of lost and gained genes**

251 Gene enrichment analysis comparing smelt with human showed that smelt lost some
252 genes involved in stimulus, reproduction, nucleotide and metal ions. Comparing with
253 salmon showed that smelt lost chemotaxis, immune system, lipid, mitochondria, neuron,
254 nucleotide and reproduction related genes (Table S11). In gained genes enrichment
255 analysis, compared with human, smelt gained chemotaxis, heparin production, immune
256 system and stimulation related genes. Compared with zebrafish and salmon, smelt gained
257 chemotaxis and immune system related genes, respectively (Table S12).

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

260 To identify differences in gene expression in smelt liver and muscle tissues at different
261 temperatures, we constructed and sequenced twenty cDNA libraries on Illumina
262 NovaSeq6000 platform with pair-end sequencing. The clean reads were generated by
263 filtering the raw reads from the NT and HT groups. Table S13 and S14 shows the
264 statistics of clean reads filtered from raw reads, and the mean quality is more than 30%
265 for each sample.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

266 **Gene expression profiling**

267 A total of 297 and 331 genes were significantly differentially expressed in female liver
268 and muscle tissues in the HT group compared to those in the NT group, respectively
269 (|log2[Fold change]| > 1, FDR < 0.05). In the female liver tissue, 297 DEGs, including
270 154 up-regulated DEGs and 143 down-regulated DEGs, were identified. In the female
271 muscle tissue, 331 DEGs, including 192 up-regulated DEGs and 139 down-regulated
272 DEGs, were identified. In the male liver tissue, 8 DEGs, including 4 up-regulated DEGs
273 and 4 down-regulated DEGs, were identified. In the male muscle tissue, 29 DEGs,
274 including 9 up-regulated and 20 down-regulated DEGs, were identified (Figure 2). The
275 volcano plot showed the DEG analysis in male liver and muscle tissues. Since the
276 number of DEGs in male tissue is significantly less than that in female tissue, the
277 following gene functional enrichment analysis only involves female tissue. Table S16
278 shows all DEGs.

279 **GO analysis of DEGs in the liver and muscle**

280 In enrichment analysis, DEGs were enriched GO terms (threshold as p < 0.05 and odds
281 ratio > 1). For the liver, 128 up-regulated and 115 down-regulated DEGs were annotated
282 with GO terms. Up-regulated GO terms including lipid metabolism, cell death, DNA
283 binding and stress related genes and down-regulated GO terms including immune system,
284 membrane protein and digestion related genes (Table 3). For the muscle, 161 up-
285 regulated and 114 down-regulated DEGs were annotated with GO terms. Up-regulated
286 GO terms including apoptosis, hypoxia, ion transportation, transcription and component
287 organization related genes and down-regulated GO terms including nicotinamide
288 nucleotide and ubiquitination related genes (Table 4).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

289 **KEGG pathway enrichment analysis of DEGs in the liver and muscle**

290 Four KEGG pathways were enriched in the liver tissue. One pathway (FDR < 0.1),
291 including the terpenoid backbone biosynthesis pathway, was up-regulated, while three
292 pathways (FDR < 0.05), including influenza A, Jak-STAT signaling, and pancreatic
293 secretion pathways, were down-regulated. Thirty-two KEGG pathways were enriched in
294 the muscle tissue. Mitogen-activated protein kinase (MAPK) signaling (FDR < 0.05),
295 oxytocin signaling pathways (FDR < 0.05) and HIF-1 signaling pathway (FDR < 0.05)
296 were up-regulated, while four pathways (FDR < 0.05), including inflammatory bowel
297 disease (IBD), Leishmaniasis, African trypanosomiasis, and type I diabetes mellitus
298 pathways, were down-regulated (Figure 3).

299 Discussion

300 *H. nipponensis* is one of popular freshwater fish species in South Korean winter. With the
301 rise in global warming, fish are suffering from heat stress, and a large number of deaths
302 have been observed, resulting in great economic losses to the aquaculture industry. The
303 present study reports the first high-quality draft genome assembly of *H. nipponensis* from
304 the family Osmeridae (order Osmeriformes) and the analysis of transcriptome in heat
305 stress to elucidate the mechanism of cellular response in liver and muscle tissues.

306 **Genome sequencing**

307 We successfully generated *H. nipponensis* draft genome assembly of 1,987 highly
308 contiguous contigs (498.3 Mbp), with a high N_{50} (0.47 Mbp). The estimated genome size
309 of 464 Mbp was consistent with that of another member of Osmeridae and greater than
310 that of *Osmerus eperlanus* (European smelt, 342.8 Mbp) in the family Osmeridae

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

311 (Malmstrøm et al., 2016). Compared with other species (mammal, bird, bony fish and
312 cartilaginous fish), the genome size of smelt is smaller than that of other species, but the
313 GC content of genome and coding sequence (CDS) and GC content of CDS are similar to
314 those of other species, and the number of CDS is even more than that of other species
315 (Table S15). The genome size of *H. nipponensis* is within the range of most published
316 fish genome size(Fan et al., 2020).

317 **Lost and gained genes**

318 In the lost gene analysis, we compared *H. nipponensis* to human (*H. sapiens*), Zebrafish
319 (*D. rerio*), and Atlantic salmon (*S. salar*). The lowest number of lost genes were observed
320 when comparing *H. nipponensis* to *S. salar*, because salmon was a cold-water species that
321 lived in similar water temperature. In addition, these two species are both anadromous,
322 also, Osmeriformes are close relatives of the Salmoniformes. One of smelt species that
323 rainbow smelt (*Osmerus mordax*) has high levels of similarity (86%) to salmonid
324 genes(Von Schalburg et al., 2008). In the lost gene enrichment analysis, compared to
325 human, smelt may not feel bitter taste, because smelt eat phytoplankton, they may lose
326 their taste buds of bitterness in order to survive. In study, comparing naked mole rat
327 (NMR) and human genome, NMR lost their receptors for bitter taste, because NMR live
328 underground and their staple food is plant roots that are most of bitter taste, which
329 facilitates this evolution (Kim et al., 2011). Moreover, the spermatogenesis of smelt may
330 be different from that of human. With evolution, the genes involved in sex- and
331 reproduction-related genes, such as mating behavior, fertilization, spermatogenesis and
332 sex determination evolve with environmental changes(Volff, 2005). The function of
333 cytosine is less than that of human, ammonia may not be discharged smoothly when

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

334 nitrogen is excreted. Compared with salmon, sperm capacitation of smelt may be
335 different from that of salmon. In gained genes functional enrichment analysis, compared
336 with human, smelt have more genes related to heparin production, which may be due to
337 the fact that living environment of smelt is low temperature, and heparin may prevent
338 blood coagulation. Moreover, smelt may be more sensitive to light than human. Fish live
339 in a different light environment from terrestrial species. However, water absorbs light, so
340 as the water depth increases, the amount of light available decreases. Compared with
341 zebrafish and salmon, the immune system of smelt is more complex.

342 **Transcriptome analysis**

343 The comparative gene expression profiles in heat-stressed and normal groups are useful
344 for understanding the mechanism of cellular responses under heat stress. In the HT group,
345 when the temperature reached 23°C, smelt began to lose the population equilibrium,
346 beginning to die, and there was no dead fish in the NT group. The liver is a vital
347 metabolic organ that relates to stress response. In the gene enrichment analysis, we found
348 that some genes involved in lipid metabolism, like lipid biosynthetic process, isoprenoid
349 metabolic process and steroid metabolic process. Lipid is the main component of cell
350 membrane, because acute heat stress may increase the fluidity of cell membrane, and the
351 increase of lipid metabolism may be to maintain the stability of cell membrane. Also,
352 lipid is the main component of fat, which can block the external heat, and lipid
353 metabolism may increase the synthesis of fat to block the external heat. In *T. bernacchii*
354 thermal acclimation study, there was resulted in an increase of membrane saturated fatty
355 acids(Malekar et al., 2018). Moreover, lipid might help to maintain energy homeostasis in
356 fish, as lipids were the basic components of sterols. Due to the membrane unique

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

357 molecular structure, it has a thermal sensitive macromolecular structure. Environmental
358 stress activates lipid metabolic enzymes and targets downstream signaling
359 pathway(Balogh et al., 2013). Poikilothermic organisms are able to maintain their
360 membrane fluidity for temperature fluctuation-induced cellular disturbance by regulating
361 the composition of membrane lipids through physiological and biochemical mechanisms
362 of homeoviscous adaptation(Mendoza, 2014). After heat shock, the plasma lipid peroxide
363 level increase gradually and severe heat stress affects the redox state and causes oxidative
364 stress in salmon(Nakano et al., 2014). When fish suffer from hypoxia, fat metabolism is
365 enhanced. Isoprenoid and terpenoid metabolic process-related genes such as *HMGCS1*,
366 *HMGCR*, and *FDPS* were up-regulated in liver tissue. These genes are involved in
367 squalene synthesis. In human cancer cells study, hypoxic cells display profound
368 accumulation of squalene(Kucharzewska, Christianson, & Belting, 2015). Additionally, it
369 plays an anti-oxidant role that eliminates ROS produced under stress(Micera et al., 2020).
370 Squalene synthase is up-regulated in the liver. This was the main material for the
371 synthesis of sterols, and steroid metabolic process-related genes were up-regulated in the
372 liver under heat stress. In other studies, heat stress led to the increase in estrogen level,
373 and hormonal disorders led to an imbalance in the number of males and females in the
374 population (Shi et al., 2019). Some genes related to regulation of apoptotic process were
375 significantly up-regulated under heat stress. Although cells have different protective
376 mechanisms under stress, the enhancement of stress can lead to cell signal interruption,
377 extensive DNA damage, and cell apoptosis (Cheng et al., 2015). DNA damage is caused
378 by hypoxia, leading to replication stress. Moreover, some genes related to regulation of
379 cellular response to stress such as *PLRG1*, *RTEL1*, and *ING2* were up-regulated. *PLRG1*

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

380 is involved in pre-mRNA splicing as a component of the spliceosome critical for heat
381 environment adaptation (Huang et al., 2018). *RTEL1* and *ING2* play a role in DNA repair
382 under heat stress, indicating that liver tissue has replication problems under heat stress.
383 However, some genes related to erythrocyte characteristics, including GO term
384 erythrocyte differentiation, erythrocyte development, and regulation of anatomical
385 structure size, were significantly down-regulated under heat stress. Erythrocytes are
386 cellular mediators of the immune response in teleost fish, thus, the immunity of smelt was
387 reduced under heat stress. In addition, the genes related to cell membrane signal
388 transduction were down-regulated in heat stress group, which indicates that it is difficult
389 for cells to communicate with extracellular under the condition of homeostasis imbalance.
390 Moreover, digestion related genes have also been down-regulated, suggesting that
391 abnormal temperatures lead to feeding may be reduced, so fish should be fed less in hot
392 summer. Study has shown that the digestive enzymes activity was significantly affected
393 by abnormal temperature(Pimentel et al., 2015). The effect of water temperature on
394 digestive enzymes of fish depends on species, because the optimal temperature of
395 enzyme activity is usually in the temperature range corresponding to the fish
396 habitat(Volkoff & Rønnestad, 2020). The muscle is greatly affected by heat stress, as, in
397 most species, this tissue constitutes approximately 50% of the body mass. Under heat
398 stress, apoptosis, hypoxia, ion transportation, transcription- and component organization-
399 related genes were up-regulated. Due to high temperature, muscle cells lose calcium
400 balance, which will cause muscle spasm, and muscle contraction-related genes that those
401 actin-filament genes were up-regulated, also, hypoxia can cause muscle contraction. In
402 oxygen consumption study, when the water temperature rises from 14°C to 20°C, the

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

403 oxygen consumption of Delta smelt increases with the increase of temperature(Jeffries et
404 al., 2016). In addition, nicotinamide adenine dinucleotide (NADH)- and ubiquitination-
405 related genes associated with cell respiration and energy metabolism were down-
406 regulated. This might be due to the decrease in collective energy metabolism caused by
407 hypoxia.

408 In pathway enrichment analysis, under heat stress, terpenoid backbone biosynthesis
409 pathway in the liver was up-regulated. The genes involved in this pathway are implicated
410 in the final synthesis of squalene, which is an important enzyme under hypoxia condition.
411 Immune-related pathways, including influenza A and Jak-STAT signaling pathways,
412 were down-regulated. Digestion-related pancreatic secretion, protein digestion, and
413 absorption pathways were down-regulated in the liver. This suggested that the digestive
414 function of fish was disrupted under heat stress. In the muscle, several genes involved in
415 MAPK signaling pathway were up-regulated under heat stress. A recent study has shown
416 that MAPK signaling pathway is activated in response to ER stress (Darling & Cook,
417 2014). The disturbance of ER environment such as the decrease in Ca^{2+} concentration or
418 the change in redox state can affect protein folding and processing. Once misfolded
419 proteins accumulate, ER stress activates a series of corresponding pathways. Under heat
420 stress, some enriched genes in oxytocin signaling pathway were also up-regulated.
421 Oxytocin activates the signal pathways of mRNA translation during ER stress (Klein et
422 al., 2016). HIF-1 signaling pathway was also up-regulated in the muscle tissue, indicating
423 that smelt suffered hypoxia stress in acute increasing temperature. *HIF1A* is the main
424 regulator of hypoxia-induced gene expression. The metabolism of smelt increases with
425 the rise in temperature, which may lead to lack of oxygen. Smelt lives in low-temperature

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

426 environments, and the oxygen solubility in low-temperature water is higher than that in
427 high-temperature water, thus, smelt is likely to have fewer red blood cells than those of
428 temperate fish. Hypoxic stress induced unfolded protein response (UPR) and autophagy.
429 Hypoxia causes perturbations in the ER activity, resulting in UPR activation. These up-
430 regulated pathways suggested that smelt suffered hypoxia and ER stress under heat stress.
431 ER stress-related genes *NFE2L1* and *ERGIC1* were up-regulated in the muscle tissue. ER
432 stress was also observed under heat stress in Atlantic salmon (Shi et al., 2019). IBD,
433 Leishmaniasis, African trypaosomiasis and type I diabetes mellitus pathways were down-
434 regulated in muscle under heat stress, these are related to immune and infectious diseases,
435 considering that heat stress has a negative effect on the inner immune system(Lyu et al.,
436 2018).

437

438 **Conclusions**

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

446

447 **Acknowledgments**

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

448 This research was supported by a grant from the Collaborative Genome Program of the
449 Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the
450 Ministry of Oceans and Fisheries (MOF) (No. 20180430) and the National Institute of
451 Fisheries Scieence (R2020034). Biao Xuan was supported by the BK21 Plus Program
452 from Ministry of Education.

453 **Literature Cited**

454 Balogh, G., Péter, M., Glatz, A., Gombos, I., Török, Z., Horváth, I., … Vígh, L. (2013).
455 Key role of lipids in heat stress management. *FEBS Letters*, *587*(13), 1970–1980.
456 https://doi.org/10.1016/j.febslet.2013.05.016

457 Basu, N., Basu, N., Todgham, A. E., Todgham, A. E., Ackerman, P. A., Ackerman, P.
458 A., … Iwama, G. K. (2002). Heat shock protein genes and their functional
459 significance in fish. *Gene*, Vol. 295, pp. 173–183. Retrieved from
460 papers2://publication/uuid/BD714DF2-29DF-4FED-8878-
461 0752FA4DC4AA%5Cnpapers2://publication/uuid/B619098C-2E68-4688-877E-
462 0C61FBDDE551

463 Cheng, C. H., Yang, F. F., Liao, S. A., Miao, Y. T., Ye, C. X., Wang, A. L., … Chen, X.
464 Y. (2015). High temperature induces apoptosis and oxidative stress in pufferfish
465 (Takifugu obscurus) blood cells. *Journal of Thermal Biology*, *53*, 172–179.
466 https://doi.org/10.1016/j.jtherbio.2015.08.002

467 Choi, S., & Kim, E. B. (2019). Complete mitochondrial genome sequence and SNPs of
468 the Korean smelt Hypomesus nipponensis (Osmeriformes, Osmeridae).
469 *Mitochondrial DNA Part B: Resources*, *4*(1), 1844–1845.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

470 https://doi.org/10.1080/23802359.2019.1613178

471 Darling, N. J., & Cook, S. J. (2014). The role of MAPK signalling pathways in the
472 response to endoplasmic reticulum stress. *Biochimica et Biophysica Acta -*
473 *Molecular Cell Research*, *1843*(10), 2150–2163.
474 https://doi.org/10.1016/j.bbamcr.2014.01.009

475 Fan, G., Song, Y., Yang, L., Huang, X., Zhang, S., Zhang, M., … He, S. (2020). Initial
476 data release and announcement of the 10,000 Fish Genomes Project (Fish10K).
477 *GigaScience*, Vol. 9. https://doi.org/10.1093/gigascience/giaa080

478 Fulda, S., Gorman, A. M., Hori, O., & Samali, A. (2010). Cellular stress responses: Cell
479 survival and cell death. *International Journal of Cell Biology*.
480 https://doi.org/10.1155/2010/214074

481 Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., …
482 Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using
483 the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8),
484 1494–1512. https://doi.org/10.1038/nprot.2013.084

485 Huang, J., Li, Y., Liu, Z., Kang, Y., & Wang, J. (2018). Transcriptomic responses to heat
486 stress in rainbow trout Oncorhynchus mykiss head kidney. *Fish and Shellfish*
487 *Immunology*, *82*(April), 32–40. https://doi.org/10.1016/j.fsi.2018.08.002

488 Jeffries, K. M., Connon, R. E., Davis, B. E., Komoroske, L. M., Britton, M. T., Sommer,
489 T., … Fangue, N. A. (2016). Effects of high temperatures on threatened estuarine
490 fishes during periods of extreme drought. *Journal of Experimental Biology*, Vol. 219,
491 pp. 1705–1716. https://doi.org/10.1242/jeb.134528

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

492 Kelly, T., Johnsen, H., Burgerhout, E., Tveiten, H., Thesslund, T., Andersen, Ø., &
493 Robinson, N. (2020). Low oxygen stress during early development influences
494 regulation of hypoxia-response genes in farmed atlantic salmon (Salmo salar). *G3:*
495 *Genes, Genomes, Genetics*, *10*(9), 3179–3188.
496 https://doi.org/10.1534/g3.120.401459

497 Kim, E. B., Fang, X., Fushan, A. A., Huang, Z., Lobanov, A. V., Han, L., … Gladyshev,
498 V. N. (2011). Genome sequencing reveals insights into physiology and longevity of
499 the naked mole rat. *Nature*, *479*(7372), 223–227.
500 https://doi.org/10.1038/nature10533

501 Klein, B. Y., Tamir, H., Hirschberg, D. L., Ludwig, R. J., Glickstein, S. B., Myers, M. M.,
502 & Welch, M. G. (2016). Oxytocin opposes effects of bacterial endotoxin on ER-
503 stress signaling in Caco2BB gut cells. *Biochimica et Biophysica Acta - General*
504 *Subjects*, *1860*(2), 402–411. https://doi.org/10.1016/j.bbagen.2015.10.025

505 Kucharzewska, P., Christianson, H. C., & Belting, M. (2015). Global profiling of
506 metabolic adaptation to hypoxic stress in human glioblastoma cells. *PLoS ONE*, Vol.
507 10. https://doi.org/10.1371/journal.pone.0116740

508 Logan, C. A., & Buckley, B. A. (2015). Transcriptomic responses to environmental
509 temperature in eurythermal and stenothermal fishes. *Journal of Experimental*
510 *Biology*, *218*(12), 1915–1924. https://doi.org/10.1242/jeb.114397

511 Lu, Y., Wu, Z., Song, Z., Xiao, P., Liu, Y., Zhang, P., & You, F. (2016). Insight into the
512 heat resistance of fish via blood: Effects of heat stress on metabolism, oxidative
513 stress and antioxidant response of olive flounder Paralichthys olivaceus and turbot

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

514 Scophthalmus maximus. *Fish and Shellfish Immunology*, *58*, 125–135.
515 https://doi.org/10.1016/j.fsi.2016.09.008

516 Lyu, L., Wen, H., Li, Y., Li, J., Zhao, J., Zhang, S., … Wang, X. (2018). Deep
517 Transcriptomic Analysis of Black Rockfish (Sebastes schlegelii) Provides New
518 Insights on Responses to Acute Temperature Stress. *Scientific Reports*, Vol. 8.
519 https://doi.org/10.1038/s41598-018-27013-z

520 Malekar, V. C., Morton, J. D., Hider, R. N., Cruickshank, R. H., Hodge, S., & Metcalf, V.
521 J. (2018). Effect of elevated temperature on membrane lipid saturation in Antarctic
522 notothenioid fish. *PeerJ*, *2018*(5), 1–25. https://doi.org/10.7717/peerj.4765

523 Malmstrøm, M., Matschiner, M., Tørresen, O. K., Star, B., Snipen, L. G., Hansen, T.
524 F., … Jentoft, S. (2016). Evolution of the immune system influences speciation rates
525 in teleost fishes. *Nature Genetics*, *48*(10), 1204–1210.
526 https://doi.org/10.1038/ng.3645

527 Mendoza, D. De. (2014). Temperature sensing by membranes. *Annual Review of*
528 *Microbiology*, *68*(April), 101–116. https://doi.org/10.1146/annurev-micro-091313-
529 103612

530 Micera, M., Botto, A., Geddo, F., Antoniotti, S., Bertea, C. M., Levi, R., … Querio, G.
531 (2020). Squalene: More than a step toward sterols. *Antioxidants*, Vol. 9, pp. 1–14.
532 https://doi.org/10.3390/antiox9080688

533 Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: An
534 automatic genome annotation and pathway reconstruction server. *Nucleic Acids*
535 *Research*, *35*(SUPPL.2), 182–185. https://doi.org/10.1093/nar/gkm321

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

536 Nakano, T., Kameda, M., Shoji, Y., Hayashi, S., Yamaguchi, T., & Sato, M. (2014).
537 Effect of severe environmental thermal stress on redox state in salmon. *Redox*
538 *Biology*, *2*(1), 772–776. https://doi.org/10.1016/j.redox.2014.05.007

539 Narum, S. R., & Campbell, N. R. (2015). Transcriptomic response to heat stress among
540 ecologically divergent populations of redband trout. *BMC Genomics*, Vol. 16.
541 https://doi.org/10.1186/s12864-015-1246-5

542 Olsvik, P. A., Vikeså, V., Lie, K. K., & Hevrøy, E. M. (2013). Transcriptional responses
543 to temperature and low oxygen stress in Atlantic salmon studied with next-
544 generation sequencing technology. *BMC Genomics*, *14*(1).
545 https://doi.org/10.1186/1471-2164-14-817

546 Pimentel, M. S., Faleiro, F., Diniz, M., Machado, J., Pousão-Ferreira, P., Peck, M. A., …
547 Rosa, R. (2015). Oxidative stress and digestive enzyme activity of flatfish larvae in a
548 changing ocean. *PLoS ONE*, Vol. 10. https://doi.org/10.1371/journal.pone.0134082

549 Qiang, J., Bao, W. J., Tao, F. Y., He, J., Li, X. H., Xu, P., & Sun, L. Y. (2017). The
550 expression profiles of miRNA-mRNA of early response in genetically improved
551 farmed tilapia (Oreochromis niloticus) liver by acute heat stress. *Scientific Reports*,
552 *7*(1), 1–15. https://doi.org/10.1038/s41598-017-09264-4

553 Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor
554 package for differential expression analysis of digital gene expression data.
555 *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

556 Saruwatari, T., Lopez, J. A., & Pietsch, T. W. (1997). A Revision of the Osmerid Genus
557 Hypomesus Gill (Teleostei : Salmoniformes), with the Description of a New Species

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

558 from the Southern Kuril IslandsA Revision of the Osmerid Genus Hypomesus Gill
559 (Teleostei : Salmoniformes), with the Description of a New Spec. *Species Diversity*,
560 Vol. 2, pp. 59–82. https://doi.org/10.12782/specdiv.2.59

561 Shi, K. P., Dong, S. L., Zhou, Y. G., Li, Y., Gao, Q. F., & Sun, D. J. (2019). RNA-seq
562 reveals temporal differences in the transcriptome response to acute heat stress in the
563 Atlantic salmon (Salmo salar). *Comparative Biochemistry and Physiology - Part D:*
564 *Genomics and Proteomics*, *30*(February), 169–178.
565 https://doi.org/10.1016/j.cbd.2018.12.011

566 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M.
567 (2015). BUSCO: Assessing genome assembly and annotation completeness with
568 single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212.
569 https://doi.org/10.1093/bioinformatics/btv351

570 Volff, J. N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity*, Vol. 94,
571 pp. 280–294. https://doi.org/10.1038/sj.hdy.6800635

572 Volkoff, H., & Rønnestad, I. (2020). Effects of temperature on feeding and digestive
573 processes in fish. *Temperature*, *0*(0), 1.
574 https://doi.org/10.1080/23328940.2020.1765950

575 Von Schalburg, K. R., Leong, J., Cooper, G. A., Robb, A., Beetz-Sargent, M. R., Lieph,
576 R., … Koop, B. F. (2008). Rainbow smelt (Osmerus mordax) genomic library and
577 EST resources. *Marine Biotechnology*, Vol. 10, pp. 487–491.
578 https://doi.org/10.1007/s10126-008-9089-6

579 Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., … Earl, A.

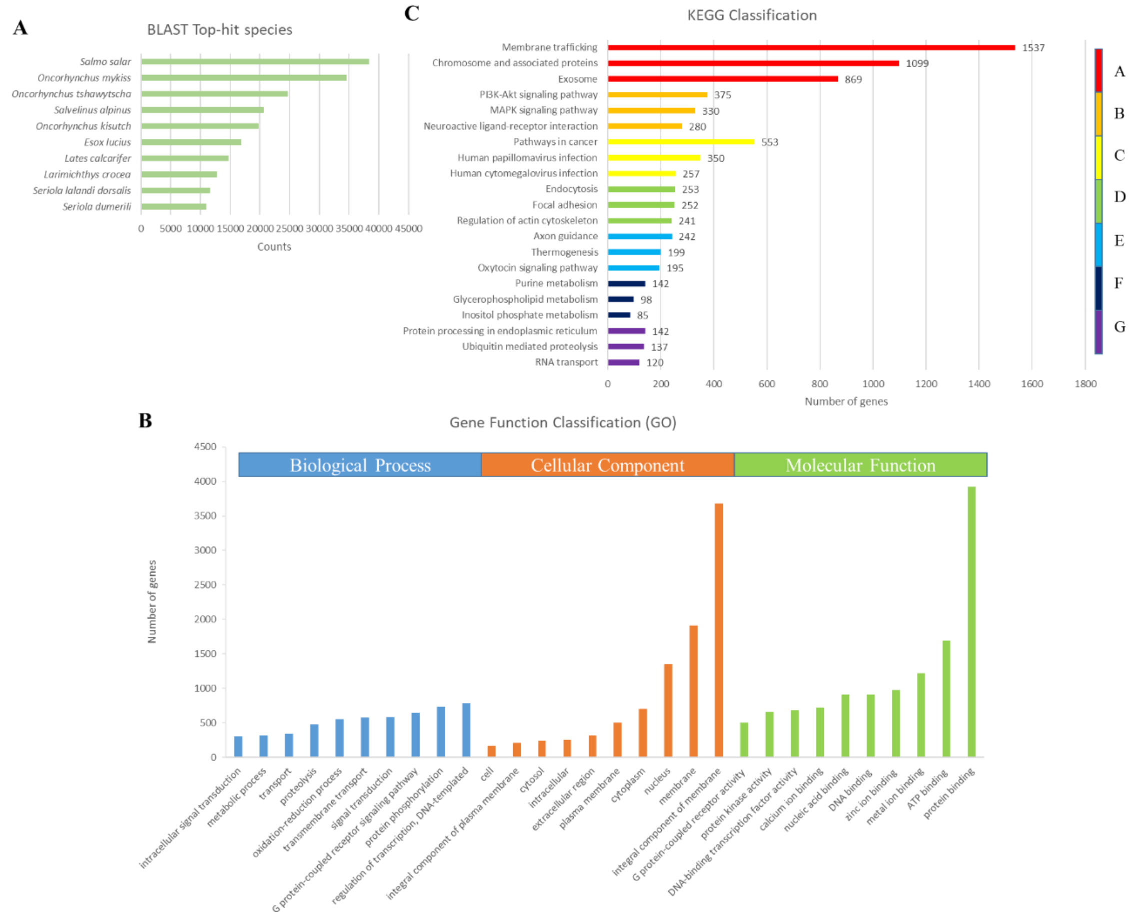
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

580 M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection
581 and genome assembly improvement. *PLoS ONE*, Vol. 9.
582 https://doi.org/10.1371/journal.pone.0112963

583 Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for
584 comparing biological themes among gene clusters. *OMICS A Journal of Integrative*
585 *Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

586
587
588
589
590
591
592
593
594
595
596
597
598

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



599

600 **Figure 1 Annotation and functional classification of unigenes in genome of *H.***

601 ***nipponensis*.** (A) Blast top 10 hit species. (B) GO analysis was performed at level 2 for

602 the three main categories (biological process, cellular component and molecular function).

603 (C) Pathway assignment based on the Kyoto Encyclopedia of Genes and Genomes

604 (KEGG) database. Unigenes were classified into seven main categories (A: Genetic

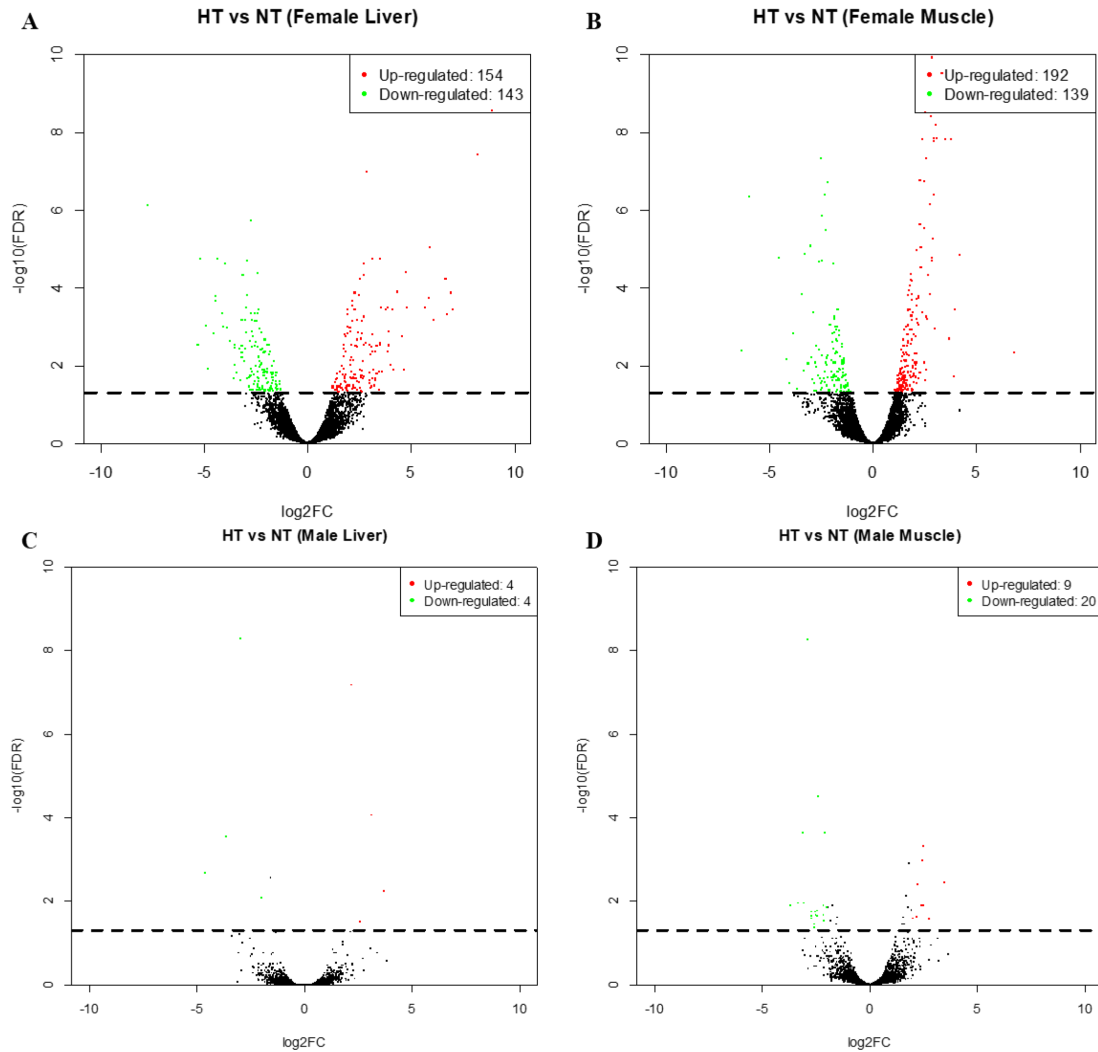
605 Information Processing; B: Metabolism; C: Organismal Systems; D: Cellular Processes;

606 E: Human Diseases; F: Environmental Information Processing; G: Brite Hierarchies).

607

608

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



609

610 **Figure 2. Differential expression of high-temperature in *H. nipponensis* versus low-**
611 **temperature.** A. DEGs in female liver. B. DEGs in female muscle. C. DEGs in male
612 liver. D. DEGs in male muscle. The dots above the dotted line are FDR < 0.05, |log2FC| ≥
613 1.

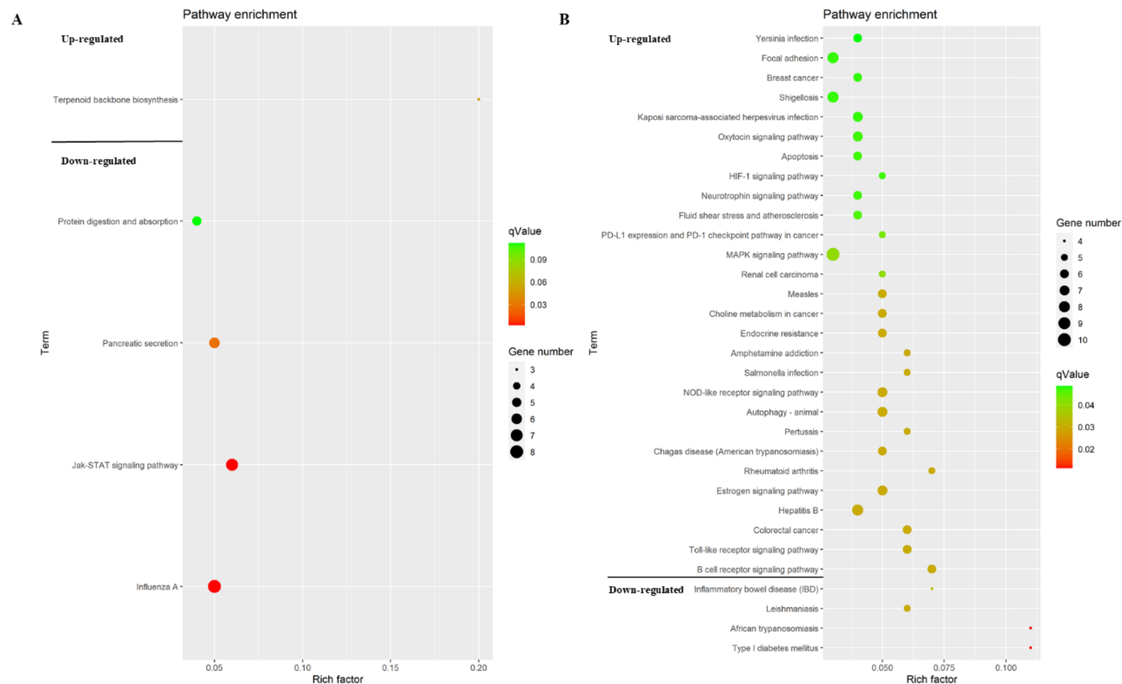
614

615

616

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

617



618

619 **Figure 3 An overview of the KEGG pathway significantly enriched in DEGs in the**

620 **(A) Liver, (B) Muscle.** The specific pathways are plotted along the y-axis, and the x-axis

621 indicates the enrichment factor. The size colored dots indicates the number of

622 significantly DEGs associated with each corresponding pathway: pathway with larger-

623 sized dots contain a higher number of genes. The color of the each dot indicates the

624 corrected p-value for the corresponding pathway.

625

626

627

628

629

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. *Hypomesus nipponensis* draft genome statistics.

Index	*Hypomesus nipponensis*
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 N50	464,523
GC contents	45.63%
BUSCO (Actinopterygii) complete	93.58%
BUSCO (Vertebrata) complete	95.32%
CEGMA complete	92.74%

630

Table 2. *Hypomesus 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	172,480
	Average intron length	1,245.81
	Total intron length	214,877,748
	Genome coverage	43.07%

631

632

633

634

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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
Up-regulated	Lipid metabolism	GO:0008610	lipid biosynthetic process	15955	128	127	7	0.00	6.87
		GO:0006629	lipid metabolic process	15955	128	371	10	0.00	3.36
		GO:0006720	isoprenoid metabolic process	15955	128	24	4	0.00	20.76
		GO:0008299	isoprenoid biosynthetic process	15955	128	13	3	0.00	28.74
		GO:0006721	terpenoid metabolic process	15955	128	16	2	0.01	15.57
		GO:0008202	steroid metabolic process	15955	128	35	4	0.00	14.24
		GO:0016125	sterol metabolic process	15955	128	17	3	0.00	21.98
		GO:0006694	steroid biosynthetic process	15955	128	21	3	0.00	17.79
	Cell death	GO:0042981	regulation of apoptotic process	15955	128	154	6	0.00	4.86
		GO:0043067	regulation of programmed cell death	15955	128	157	6	0.00	4.76
		GO:0010941	regulation of cell death	15955	128	164	6	0.00	4.56
	DNA binding	GO:0030983	mismatched DNA binding	15955	128	13	2	0.01	19.16
		GO:0006298	mismatch repair	15955	128	15	2	0.01	16.61
		GO:0003690	double-stranded DNA binding	15955	128	117	4	0.02	4.26
	Stress	GO:0080135	regulation of cellular response to stress	15955	128	49	3	0.01	7.63
		GO:0080134	regulation of response to stress	15955	128	78	3	0.03	4.79
Down-regulated	Immune system	GO:0030099	myeloid cell differentiation	15955	115	44	4	0.00	12.60
		GO:0030218	erythrocyte differentiation	15955	115	23	3	0.00	18.08
		GO:0048821	erythrocyte development	15955	115	10	2	0.00	27.69
		GO:0061515	myeloid cell development	15955	115	16	2	0.01	17.33
	Membrane protein	GO:0005834	heterotrimeric G-protein complex	15955	115	30	3	0.00	13.87
		GO:1905360	GTPase complex	15955	115	30	3	0.00	13.87
	Digestion	GO:0007586	digestion	15955	115	13	2	0.01	21.33

635

636

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437215; this version posted March 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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
Up-regulated	Apoptosis	GO:0097190	apoptotic signaling pathway	15955	161	16	3	0.00	18.56
		GO:0008630	intrinsic apoptotic signaling pathway in response to DNA damage	15955	161	8	2	0.00	24.74
		GO:0097193	intrinsic apoptotic signaling pathway	15955	161	13	2	0.01	15.24
	Hypoxia	GO:0036293	response to decreased oxygen levels	15955	161	17	2	0.02	11.65
		GO:0001666	response to hypoxia	15955	161	17	2	0.02	11.65
	Ion transportation	GO:0015278	calcium-release channel activity	15955	161	30	3	0.00	9.91
		GO:0099604	ligand-gated calcium channel activity	15955	161	31	3	0.00	9.59
		GO:0019722	calcium-mediated signaling	15955	161	15	2	0.01	13.20
	Transcription	GO:0003700	DNA-binding transcription factor activity	15955	161	525	17	0.00	3.21
		GO:0140110	transcription regulator activity	15955	161	608	18	0.00	2.93
		GO:0000981	DNA-binding transcription factor activity, RNA polymerase II-specific	15955	161	102	5	0.00	4.86
		GO:0006355	regulation of transcription, DNA-templated	15955	161	1038	21	0.00	2.00
	Component organization	GO:0051639	actin filament network formation	15955	161	2	2	0.00	98.80
		GO:0032432	actin filament bundle	15955	161	4	2	0.00	49.42
		GO:0051017	actin filament bundle assembly	15955	161	22	3	0.00	13.50
		GO:0061572	actin filament bundle organization	15955	161	24	3	0.00	12.38
Down-regulated	Nicotinamide nucleotide	GO:0019359	nicotinamide nucleotide biosynthetic process	15955	114	13	2	0.01	21.52
		GO:0046496	nicotinamide nucleotide metabolic process	15955	114	15	2	0.01	18.65
		GO:0019363	pyridine nucleotide biosynthetic process	15955	114	15	3	0.00	27.93
		GO:0019362	pyridine nucleotide metabolic process	15955	114	16	3	0.00	26.19
		GO:0072525	pyridine-containing compound biosynthetic process	15955	114	18	3	0.00	23.29
		GO:0072524	pyridine-containing compound metabolic process	15955	114	19	3	0.00	22.07
	Ubiquitination	GO:0061630	ubiquitin protein ligase activity	15955	114	73	3	0.02	5.75
		GO:0061659	ubiquitin-like protein ligase activity	15955	114	73	3	0.02	5.75

637