A genome-wide association study of heat stress-associated SNPs in catfish

Y. Jin, T. Zhou, X. Geng, S. Liu, A. Chen, J. Yao, C. Jiang, S. Tan, B. Su and Z. Liu

The Fish Molecular Genetics and Biotechnology Laboratory, School of Fisheries, Aquaculture and Aquatic Sciences and Program of Cell and Molecular Biosciences, Aquatic Genomics Unit, Auburn University, Auburn, AL 36849, USA.

Summary

Heat tolerance is a complex and economically important trait for catfish genetic breeding programs. With global climate change, it is becoming an increasingly important trait. To better understand the molecular basis of heat stress, a genome-wide association study (GWAS) was carried out using the 250 K catfish SNP array with interspecific backcross progenies, which derived from crossing female channel catfish with male F1 hybrid catfish (female channel catfish × male blue catfish). Three significant associated SNPs were detected by performing an EMMAX approach for GWAS. The SNP located on linkage group 14 explained 12.1% of phenotypical variation. The other two SNPs, located on linkage group 16, explained 11.3 and 11.5% of phenotypical variation respectively. A total of 14 genes with heat stress related functions were detected within the significant associated regions. Among them, five genes—TRAF2, FBXW5, ANAPC2, UBR1 and KLHL29— have known functions in the protein degradation process through the ubiquitination pathway. Other genes related to heat stress include genes involved in protein biosynthesis (PRPF4 and SYNCRIP), protein folding (DNAJC25), molecule and iron transport (SLC25A46 and CLIC5), cytoskeletal reorganization (COL12A1) and energy metabolism (COX7A2, PLCB1 and PLCB4) processes. The results provide fundamental information about genes and pathways that is useful for further investigation into the molecular mechanisms of heat stress. The associated SNPs could be promising candidates for selecting heat-tolerant catfish lines after validating their effects on larger and various catfish populations.

Keywords climate change, fish, genome, hypoxia, QTL.
stress (Nayan et al. 2012). However, no genome-wide research on QTL for heat stress has been conducted in aquaculture species. Here, we conducted a genome-wide scan for QTLs conferring resistance to heat stress using interspecific backcross progenies and the 250 K catfish SNP array (Liu et al. 2014), with the objective of initial understanding of the genomic regions important for heat stress in catfish.

All procedures involving the handling and treatment of fish during this study were approved by the Auburn University Institutional Animal Care and Use Committee (IACUC) prior to the initiation of the experiments. All animal procedures were carried according to the Guide for the Care and Use of Laboratory Animal and the Animal Welfare Act in the United States. Catfish progenies generated from two backcross families by crossing male F1 hybrid catfish (female channel catfish x male blue catfish) with female channel catfish were used for the heat stress experiment. The two families were genetically independent from each other and underwent heat treatment separately. Each family containing 315 catfish progenies was transferred to an experimental tank for heat stress treatment after rearing for 2 weeks prior to challenge. The heat stress treatment followed the procedures conducted by Liu et al. (2013). Briefly, all fish were acclimated for 72 h at ambient temperature (24 °C) before treatment. Water temperature was increased by 1 °C/h until it reached 36 °C. Then, the temperature was held constant at 36 °C, and the fish were closely monitored for signs of stress. The first fish showing loss of equilibrium was observed after 3 h at 36 °C in both of the families, and the last fish showing loss of equilibrium was observed about 55 h thereafter. The first and last 48 fish individuals losing balance in each family, which represented the most heat sensitive group and most heat tolerant group respectively, were continuously removed from the tank and sampled. Overall, based on the selective genotyping method (Darvasi & Soller 1992), 192 catfish backcross progenies (502.9 g average body weight) were selected from the extremes of tolerance capability of the 630 fish. Blood from each sampled fish was collected for DNA isolation.

The genotyping of each DNA sample was performed using the catfish 250 K SNP array (Liu et al. 2014) at GeneSeek. The SVS (SNP & VARIATION SUITE, Version 8.1) software package was used to perform the quality filtering procedures and the following statistical analysis. Briefly, a total of 200 584 SNPs were kept in the subsequent GWAS after filtering out SNPs with an inheritance or genotyping error, a minor allele frequency <5% or a call rate <95%. Linkage disequilibrium (LD) pruning was conducted with a window size of 50 SNPs, a step of five SNPs, and r² threshold of 0.5. The number of independent SNP markers and LD blocks was 8091. With known family pedigree information, principal components analysis was conducted using eigenvalues as coordinates to visualize the sample structure (Fig. S1). Apparently, the two families were distantly related. EMMAX (Efficient Mixed-Model Association eXpedited) analysis was carried out using all informative SNPs with the first two principal component scores and the fish body weight as covariates for genome-wide association (Kang et al. 2010). The EMMAX is efficient for controlling population stratification, especially between-family genotype differences in our study (Kang et al. 2010; Geng et al. 2015). A Manhattan plot of the calculated −log₁₀(P-value) results for each SNP was generated by mapping SNPs to the catfish genome using the SVS software (Fig. 1). Threshold P-value for genome-wide significance was determined based on 5% Bonferroni correction with the estimated number of 8091 independent markers. Thus, the genome-wide significance threshold was 0.05/8091 = 6.180 × e-6 [−log₁₀ (P-value) = 5.209].

The GWAS identified three significant SNP markers conferring response to heat stress at the genome-wide significance level [−log₁₀ (P-value) > 5.209]. As shown in

![Manhattan plot of genome-wide association analysis for heat stress. Genome-wide association study identified three loci associated with heat stress in catfish. The horizontal line indicates the genome-wide significant threshold: −log₁₀ (P-value) = 5.209.](image)
Table 1: The significantly associated SNPs on linkage groups 14 and 16.

<table>
<thead>
<tr>
<th>Linkage group</th>
<th>SNP ID</th>
<th>Position</th>
<th>Allele</th>
<th>$-\log_{10} (P\text{-value})$</th>
<th>MAF</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>AX-85318076</td>
<td>14 002 635</td>
<td>C/A</td>
<td>5.688</td>
<td>0.250</td>
<td>0.121</td>
</tr>
<tr>
<td>16</td>
<td>AX-8615098</td>
<td>15 552 403</td>
<td>G/A</td>
<td>5.439</td>
<td>0.322</td>
<td>0.115</td>
</tr>
<tr>
<td>16</td>
<td>AX-85265807</td>
<td>18 651 314</td>
<td>T/C</td>
<td>5.359</td>
<td>0.153</td>
<td>0.113</td>
</tr>
</tbody>
</table>

MAF, minor allele frequency; $R^2$, SNP indicates the ratio of phenotypic variation.

Table 1, one SNP (AX-85318076) was located on linkage group 14 and the other two SNPs (AX-8615098 and AX-85265807) were located on linkage group 16. Their minor allele frequencies ranged from 0.153 to 0.322, and the ratio of phenotypic variation ($R^2$) explained by the SNPs ranged from 0.113 to 0.121.

In addition to mapping the heat-related SNPs, we sought to explore the genes in the catfish genome sequence (Liu et al. 2016) in 1-Mb windows (SNP position ± 0.5 Mb) surrounding each identified promising SNP. The nearby genes for each significant SNP were predicted from the catfish reference genome sequences (Liu et al. 2016) by using FGENESH (Solovyev et al. 2006) and GENSCAN (Burge & Karlin 1997). The predicted genes were annotated by BLAST analysis against the non-redundant protein database (Pruitt et al. 2007). The biological processes of the annotated genes were retrieved from the Uniprot database (www.uniprot.org). Finally, a total of 14 candidate genes were determined and are listed in Table S1 with their corresponding biological processes related to heat stress.

In catfish, response to heat stress has been reported to be related to several physiological and gene pathways by RNA-Seq analysis. For instance, genes involved in protein folding and degradation, protein biosynthesis, and energy metabolic process were highly induced under lethal temperatures (Liu et al. 2013). In general, the results demonstrated that complex molecular mechanisms are involved in heat tolerance other than simply the induction of a certain category of genes. Of the 14 genes identified in this study (Table S1), five were found to have known function in protein degradation process, including TNF receptor associated factor 2 (TRAF2), F-box and WD repeat-containing protein 5 (FBXW5), anaphase promoting complex subunit 2 (ANAPC2), E3 ubiquitin-protein ligase UBR1 (UBR1) and kelch-like family member 29 (KLHL29). Interestingly, all of them are involved in the protein ubiquitination pathway.

The protein ubiquitination pathway has been reported to play a crucial role in response to heat stress in catfish (Liu et al. 2013), goby fish (Logan & Somero 2010, 2011), Arctic char (Quinn et al. 2011) and bluefin tuna (Castilho et al. 2009). This may suggest that heat stress causes abnormal fold and irreversible damage to proteins, which are then unable to enter the molecular chaperone pathway (Buckley et al. 2006; Logan & Somero 2011). In order to avoid forming cytotoxic aggregates, such damaged proteins need to be removed via proteolytic degradation by covalently tagging with multiple units of ubiquitin when conjugated to a damaged polypeptide (Luo & Le 2010; Liu et al. 2013). Therefore, there is an increasing necessity of degradation for cells that are suffering sufficient levels of protein damage under such lethal heat treatment.

In addition to the need for enhanced protein degradation process, cellular response to heat stress involves a range of biological mechanisms to stabilize cellular function such as inhibition of DNA and protein synthesis, cell cycle arrest, molecule and iron transport, cytoskeleton reorganization and increased apoptosis (Fuquay 1981; Matsuki et al. 2003; Liu et al. 2013). Here, we also found genes participating in these biological processes and pathways in the genome-wide significantly associated regions (Table S1). These genes include those involved in protein biosynthesis (PRPF4 and SYNCRIP), protein folding (DNAJC25), molecule and iron transport (SLC25A46 and CLIC5), cytoskeletal reorganization (COL12A1) and energy metabolism (COX7A2, PLCB1 and PLCB4) (Table S1). In aquaculture species, heat stress has been proved to disturb cellular homeostasis and can lead to severe retardation in growth and development, or even death (Buckley et al. 2006; Castilho et al. 2009; Logan & Somero 2011). Taken together, the 14 genes detected in this study are involved in these biological processes, suggesting their importance for heat stress response in catfish.

This is the first association analysis at the whole genome level to investigate the genomic loci and genes related to heat stress in aquaculture species. The results provide a valuable base of genes and pathways to be further investigated for their possible functions in heat stress. Considering the population specificity of QTLs and the minor allele effect in association analyses, futures studies using larger or more catfish families and various catfish strains are necessary for fine mapping and accurate GWAS for heat stress analysis.

**Acknowledgements**

This project was supported by a competitive grant (no. 2015-67015-22907) from the Tools and Resources-Animal Breeding, Genetics and Genomics Program of the USDA National Institute of Food and Agriculture (NIFA), and partially supported by USDA Aquaculture Research Program Competitive Grant no. 2014-70007-22395.
References


Darvasi A. & Soller M. (1992) Selective genotyping for determina-
tion of linkage between a marker locus and a quantitative trait locus. Theoretical and Applied Genetics 85, 353–9.


Supporting information

Additional supporting information may be found online in the supporting information tab for this article:

Figure S1 Sample structure identified by PCA with the first two principal components.

Table S1 Information of regions associated with heat stress.