ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

SHORT COMMUNICATION

doi: 10.1111/age.12482

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

Catfish is one of the top agricultural commodities and pivotal to employment opportunities in rural areas of the southeastern United States (Liu *et al.* 2011). Ongoing global climate change will lead to a continuous rise in temperature, which may become a major stressor for fish living in natural or artificial systems (Ficke *et al.* 2007). Therefore, developing heat-tolerant catfish lines becomes an important goal for genetic breeding programs through the use of various approaches such as strain selection, crossbreeding and hybridization. The hybrid catfish is expected to have a higher heat tolerance performance than that of the channel catfish (*Ictalurus punctatus*) because blue catfish (*I. furcatus*) has a more southern distribution than does channel catfish (Stewart *et al.* 2014). As a species of ectotherms, catfish must undergo and adapt to seasonal temperature changes, ranging from near freezing during the winter in the North to over 36 °C in the earthen ponds during the summer in the southeastern United States (Ju *et al.* 2002; Arnold *et al.* 2013). The shift of ambient water temperature can directly influence and/or disturb a variety of physiological functions of catfish. Thus, catfish, as a major aquaculture species, can also serve as a model species for heat stress studies (Liu *et al.* 2013).

Genome-wide association studies (GWAS) allow the detection of linked QTLs in families as well as historically accumulated recombination events. The F2 generation of hybrids produced by a backcrossing design, along with highly segregated phenotypes, provides a useful system for QTL analysis (Geng *et al.* 2015). Using a reference genome, candidate genes physically close to QTLs can be detected, which is useful for understanding the underlying biology of a trait by identifying it in proximity to QTL (Cole *et al.* 2011; Dikmen *et al.* 2013). Results generated from GWAS can facilitate the selection of breeds and species resistant to heat

Address for correspondence

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.
E-mail: liuzhan@auburn.edu

Accepted for publication 30 June 2016

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 × 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 r^2 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 $-\log_{10}$ (*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 $0.05/8091 = 6.180 \times$ e-6 [$-\log_{10}$ (*P*-value) = 5.209].

The GWAS identified three significant SNP markers conferring response to heat stress at the genome-wide significance level [$-\log_{10}$ (*P*-value) > 5.209]. As shown in

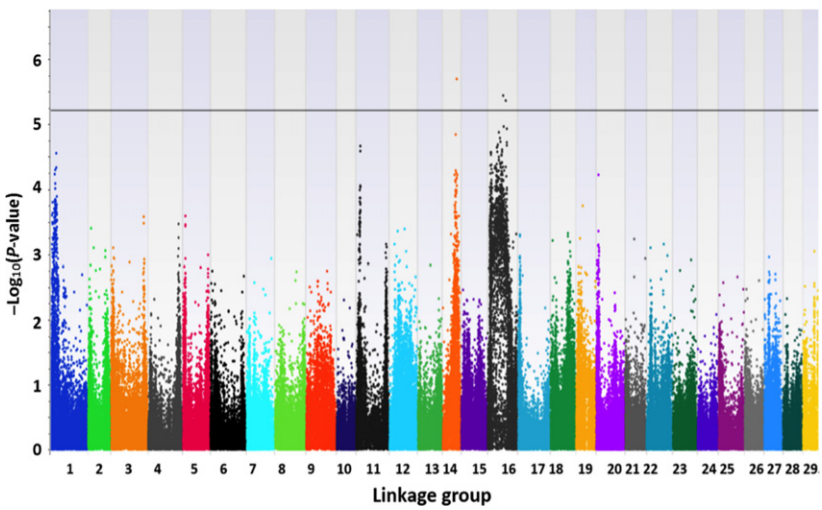

Figure 1 Manhattan plot of genome-wide association analysis for heat stress. Genome-wide association study identified three loci associated with heat stress in catfish. The horizontal line indicates the genome-wide significant threshold: $-\log_{10}$ (*P*-value) = 5.209.

Table 1 The significantly associated SNPs on linkage groups 14 and 16.

Linkage group	SNP ID	Position	Allele	$-\log_{10}$ (*P*-value)	MAF	R^2
14	AX-85318076	14 002 635	C/A	5.688	0.250	0.121
16	AX-86115098	15 552 403	G/A	5.439	0.322	0.115
16	AX-85265807	18 651 314	T/C	5.359	0.153	0.113

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-86115098 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 additional 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 charr (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 proteins, 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

Arnold M.B., Torrans E.L. & Allen P.J. (2013) Influences of cyclic, high temperatures on juvenile channel catfish growth and feeding. *North American Journal of Aquaculture* **75**, 77–84.

Buckley B.A., Gracey A.Y. & Somero G.N. (2006) The cellular response to heat stress in the goby *Gillichthys mirabilis*: a cDNA microarray and protein-level analysis. *Journal of Experimental Biology* **209**, 2660–77.

Burge C. & Karlin S. (1997) Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* **268**, 78–94.

Castilho P.C., Buckley B.A., Somero G. & Block B.A. (2009) Heterologous hybridization to a complementary DNA microarray reveals the effect of thermal acclimation in the endothermic bluefin tuna (*Thunnus orientalis*). *Molecular Ecology* **18**, 2092–102.

Cole J.B., Wiggans G.R., Ma L., Sonstegard T.S., Lawlor T.J., Crooker B.A., Van Tassell C.P., Yang J., Wang S. & Matukumalli L.K. (2011) Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC Genomics* **12**, 1.

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

Dikmen S., Cole J.B., Null D.J. & Hansen P.J. (2013) Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. *PLoS One* **8**, e69202.

Ficke A.D., Myrick C.A. & Hansen L.J. (2007) Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries* **17**, 581–613.

Fuquay J. (1981) Heat stress as it affects animal production. *Journal of Animal Science* **52**, 164–74.

Geng X., Sha J., Liu S., Bao L., Zhang J., Wang R., Yao J., Li C., Feng J. & Sun F. (2015) A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. *BMC Genomics* **16**, 196.

Ju Z., Dunham R. & Liu Z. (2002) Differential gene expression in the brain of channel catfish (*Ictalurus punctatus*) in response to cold acclimation. *Molecular Genetics and Genomics* **268**, 87–95.

Kang H.M., Sul J.H., Service S.K., Zaitlen N.A., Kong S.-Y., Freimer N.B., Sabatti C. & Eskin E. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* **42**, 348–54.

Liu S., Zhou Z., Lu J., Sun F., Wang S., Liu H., Jiang Y., Kucuktas H., Kaltenboeck L. & Peatman E. (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics* **12**, 53.

Liu S., Wang X., Sun F., Zhang J., Feng J., Liu H., Rajendran K., Sun L., Zhang Y. & Jiang Y. (2013) RNA-Seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding, and degradation in response to heat stress in catfish. *Physiological Genomics* **45**, 462–76.

Liu S., Sun L., Li Y., Sun F., Jiang Y., Zhang Y., Zhang J., Feng J., Kaltenboeck L. & Kucuktas H. (2014) Development of the catfish 250K SNP array for genome-wide association studies. *BMC Research Notes* **7**, 135.

Liu Z., Liu S., Yao J. *et al.* (2016) The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nature Communications* **7**, 11757.

Logan C.A. & Somero G.N. (2010) Transcriptional responses to thermal acclimation in the eurythermal fish *Gillichthys mirabilis* (Cooper 1864). *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **299**, R843–52.

Logan C.A. & Somero G.N. (2011) Effects of thermal acclimation on transcriptional responses to acute heat stress in the eurythermal fish *Gillichthys mirabilis* (Cooper). *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **300**, R1373–83.

Luo G.R. & Le W.D. (2010) Collective roles of molecular chaperones in protein degradation pathways associated with neurodegenerative diseases. *Current Pharmaceutical Biotechnology* **11**, 180–7.

Matsuki S., Iuchi Y., Ikeda Y., Sasagawa I., Tomita Y. & Fujii J. (2003) Suppression of cytochrome c release and apoptosis in testes with heat stress by minocycline. *Biochemical and Biophysical Research Communications* **312**, 843–9.

Nayan V., Onteru S.K. & Singh D. (2012) Genomic technologies: a way forward for learning climate resilience through cellular responses to heat stress. *Climate Resilient Livestock & Production System* **1**, 5.

Pruitt K.D., Tatusova T. & Maglott D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **35**, D61–5.

Quinn N.L., McGowan C.R., Cooper G.A., Koop B.F. & Davidson W.S. (2011) Identification of genes associated with heat tolerance in Arctic charr exposed to acute thermal stress. *Physiological Genomics* **43**, 685–96.

Solovyev V., Kosarev P., Seledsov I. & Vorobyev D. (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biology* **7**, S10.

Stewart H.A., Bosworth B., Petrie-Hanson L., Martin J.A. & Allen P.J. (2014) Effects of chronic upper temperature regimes on growth of two geographical strains of channel and hybrid catfish. *Aquaculture Research* **46**, 2407–17.

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.