Research Article

Genome-wide association study for economic traits in the large yellow croaker with different numbers of extreme phenotypes

Liang Wan[1,2Ω], Linsong Dong[2 Ω], Shijun Xiao[2], Zhaofang Han[2], Xiaoqing Wang[1*], Zhiyong Wang[1,2*]

1、 College of Animal Science and Technology, Hunan Agricultural University, Changsha, Hunan, China

2、 Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture, P.R. China, Fisheries College, Jimei University, Xiamen, Fujian, China

Ω These authors contributed equally to this work

*Correspondence: zywang@jmu.edu.cn Physical address: College of Animal Science and Technology, Hunan Agricultural University, Changsha, Hunan, China. Fax: 0592-618381. Telephone: 0592-6183816

Abstract:

A traditional genome-wide association study (GWAS) analysis genotype-phenotype associations by the vast number of genotyped individuals. This method requires large-scale samples and requires considerable sequencing costs. Extreme phenotypic sampling proposes to make GWAS more cost-efficient and applied more widely. With extreme phenotypic sampling, we performed a GWAS for *n-3* highly unsaturated fatty acids (HUFA) and eviscerated weight (EW) traits in the large yellow croaker population. Of the 32,249 and 29,748, detected SNPs for the two traits, three candidate regions were found in each trait. Three candidate regions associated with HUFA were nearby known genes on chromosomes 4 and 11, and three candidate regions were on chromosome 6 and 15 for the EW trait. By combing through our GWAS results and the biological functional analysis of the genes, we suggest that the *FABP*, *DGAT*, *ATP8B1*, *FAF2* and *CERS2* genes, as well as the *IGF2*, *BORA*, *CYP1A1*, *GRTP1*, and *HOX* genes, are promising candidate genes for n-3 HUFA and EW, respectively, in the large yellow croaker. Moreover, compared with the different numbers of the extreme phenotypic sampling, we conclude that 60% of the extreme phenotypic subsample can obtain a similar result as GWAS with whole phenotypes. Thus, extreme phenotypic sampling could save 40% of the cost for genotyping and DNA extraction without loss of the candidate regions and functional genes. Our study may provide a basis for further genomic breeding and a reference for others who want to perform GWAS with extreme phenotypes.

Introduction

The large yellow croaker (*Larimichthys crocea*) is an important maricultured fish in China with the largest yield (China Agricultural Press. 2016). The *n-3* highly unsaturated fatty acids (HUFA) content and eviscerated weight (EW) are two significant economic traits for the large yellow croaker. HUFA is also essential for the growth and development of fish. However, both of these traits could not be directly selected as phenotypes in fish breeding, so an efficient breeding technique is essential in this species and other aquacultured fish. In the traditional breeding method, genomic information was thought as a "black box", which could only be speculated but not observed. The next generation sequencing (NGS) technologies opened the "black box" and promoted the development of genomic breeding methods. GWAS has been suggested as a promising tool for genomic breeding in order to find the genetic markers associated with recorded phenotypes (Omar *et al*. 2009; Zhang *et al*. 2010), on the condition that plants and animals can be genotyped for thousands of single nucleotide polymorphisms (SNPs) at a time. Traditional GWAS analysis genotype-phenotype associations by vast number of genotyped individuals (Segura *et al*. 2012). However, this method requires large-scale samples and consumes plenty of sequencing costs, which deters some scientists from trying to solve the problem. Extreme phenotypic sampling may provide a potential solution. Studies have compared the results of extreme phenotypic sampling with that of random sampling, and they showed that extreme phenotypic sampling was more effective (Barnett *et al*. 2013; Zhou *et al*. 2016). By using extreme phenotypes, it is possible to identify the SNPs associated with specific traits (Chen *et al*. 2011; Li *et al.* 2011) or the significant genetic variants. Extreme phenotypes were also suggested to efficiently serve as the reference population to calculate estimated breeding values in genomic selection (Dong *et al*. 2016). The method of Bulk Segregation Analysis (BSA) was used to identify genotype-phenotype associations in genetically mixed pools for sequencing of extreme phenotypes. The benefit of this method is that the cost of genotyping samples from same pedigree is much cheaper than that of the traditional sampling method (Yang *et al*. 2015). Extreme phenotypic sampling provides a method to make GWAS more cost efficient and more widely applicable. There is limited economic value in every aquatic organism candidate, especially for the candidates that can only be used one time. Therefore, lower-cost methods for genotyping are required for aquatic genomic breeding. The

development of genotype-phenotype candidate regions and candidate genes offers valuable genetic resources for conservation of the large yellow croaker.

In this study, we performed GWAS for HUFA content and EW to find the main regions and candidate gene(s) associated with these two characteristics in the large yellow croaker, and different numbers of individuals with extreme phenotypes were used to evaluate the influence of GWAS. The appropriate numbers of extreme phenotypes can be a reference for the future research.

Materials and Methods

Sample preparation and Phenotypic data detection

A total of 500 and 176 large yellow croakers were used as the experimental materials in EW and n-3 HUFA studies, respectively. All fish were reared in a farm named 'Jinling Aquaculture Science and Technology Co. Ltd.' in Ning de City, Fujian Province, P.R. China. The experiment was carried out in the Key Laboratory of Healthy Mariculture for the East China Sea when the fish were 2years old. Large yellow croaker was randomly selected for the study. Two quantitative traits were measured: eviscerated weight (EW) and the content of *n-3* HUFA (Murillo *et al*. 2013). Briefly, the total lipid was extracted from the fresh muscle tissue using the chloroform methanol method. After saponification with 1 ml of 45% KOH in 15 ml ethanol, the lipid was then esterified in 85 °C for 15 min using 6.5% boron trifluoride (BF3) in methanol (Morita chemical industries Co., Ltd., Osaka, Japan). After mixing up in hexane (25 mg/ml), fatty acid methyl esters preparations were analysed by gas chromatography (GC). The temperature increase of 190–250 °C at 2 °C/min was set and helium was used as the carrier gas. The phenotypic data was descending sort, the higher and the lower phenotype values were taken as extreme phenotypes. By Shapiro-Wilk tests, EW and HUFA phenotypic data follow the normal distribution (P-value > 0.05).

DNA extraction

The dorsal fins (20-30mg) of the individual fish were collected and frozen in liquid nitrogen for subsequent DNA extraction. Total genomic DNA was prepared in 2 ml micro centrifuge tubes containing 500 μl TE buffer (120 mM NaCl, 10 mM Tris, pH 7.5, 25 mM EDTA, 0.5% SDS, and proteinase K, 0.3mg/ml). The samples were incubated at 57 °C overnight and subsequently extracted twice using the phenol and then phenol/chloroform (1:1) method. DNA was precipitated

by adding 2.5 volumes of ethanol, collected by brief centrifugation, washed twice with 75% ethanol, air dried, and dissolved in TE buffer.

Genotyping-by-sequencing (GBS) and genotyping

GBS technology was used to construct the libraries for next generation sequencing (Xiao *et al*. 2016). Genomic DNA was incubated at 37 °C with *Eco*RI and *Nla*III, 10XCutSmart™ Buffer. The restriction reactions were heat-inactivated at 60 °C by 20 min and were kept at 8 °C for the subsequent experiments. Sequencing adaptor and barcode mix, T4 DNA Ligase, 10 mM ATP and 10XCutSmart™ Buffer were incubated at 16 °C for 2 h for ligation reactions. The reactions were then heat-inactivated at 65 °C for 20 min, and the reaction systems were kept at 8°C. Then, polymerase chain reactions (PCR) experiments were performed in the reaction solutions (20 μL) containing the diluted restriction/ligation samples (4 pM, 2 μL), dNTP, Taq DNA polymerase (NEB, Ipswich, MA, USA), Illumina Primers, and Indexing Primers. The PCR procedure was 95 °C for 2 min; 15 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s; 72 °C 5 min, and kept in 4 °C. The PCR products were run on an 8% polyacrylamide gel electrophoresis. Fragments of 200~300 bp were isolated using QIAGEN QIA quick® Gel Extraction Kit and diluted for pair-end sequencing on an Illumina HiSeq 2500 sequencing platform (Illumina, Inc, San Diego, CA, USA). To avoid the negative influence of ambiguous bases for SNP detection, reads with more than 5% of N were removed. Then, the resulted reads were cleaned by the following steps: 1) discarding the reads with the quality lower than 20; 2) deleting 5bp windows in read ends with the average quality less than 20; 3) removing read pairs if one end was shorter than 50 bp. The cleaned reads were mapped to the large yellow croaker genome by BWA 0.7.6a (Li and Durbin. 2010). The mapping was preceded by a short reads alignment with BWA-MEM algorithm. The alignments were then sorted by coordinates and duplicates marked by the Sort Sam and Mark Duplicates programs, respectively, in Picard tools 1.107 (picard.sourceforge.net). The SNP markers were detected by the GATK Unified Genotype utility. The calling SNPs were discarded according to any of the following criteria (Dong *et al*. 2016): (1) missing rate ≥ 20%; (2) MAF (minor allele frequency) < 0.05; or (3) significant deviation from Hardy-Weinberg equilibrium (HWE) (p-value < 0.001). As a result, 29,748 SNPs for EW and 32,249 SNPs for n-3 HUFA were retained, and the average missing rate of markers was 11.9%. The missing SNPs were imputed by Beagle Version 3.3.2 software (Browning. 2009).

Methods for extreme phenotypes sampling

The initial phenotypes were standardized by the formula $y_i' = (y_i - u_k)/\sigma$, where u_k was the mean of phenotypes for the *kth* sex (k = 1 for male and 2 for female), and σ was the standard deviation of phenotypes after subtraction of u_k for all individuals. All experimental samples were divided into five groups in the two traits, and each group contained different numbers of extreme phenotypes (shown in Table 1).

Statistics

A single marker analysis method was used to find SNPs significantly associated with the traits. To reduce the false positive rates, we performed GWAS using the EMMAX software (Kang *et al*. 2010) in this study, which can correct for population structure and genetic relatedness (Kang *et al*. 2008). In the implementation of EMMAX, the genomic relationship matrix was first calculated using genome-wide SNPs, and then a mixed linear model was used to detect the association between a SNP and the phenotypes:

$$y_i = u + Z_{ij}\beta_j + R_i$$

where y_i is the standardized phenotypic value of an individual i, u is an overall mean, Z_{ij} is the genotype for the jth locus of individual i, β_j is regression coefficient of the jth locus, and Var(**R**) = $\mathbf{G}\sigma_g^2 + \mathbf{I}\sigma_e^2$, where **G** is the genomic relationship matrix estimated by EMMAX software, σ_g^2 is the total genetic variance, and σ_e^2 is the environmental variance. Manhattan plots were drawn by using R-package 'qqman', version 0.1.2 (https://cran.r-project.org/web/packages/qqman/).

To identify the biological functions of nearby genes associated with HUFA and EW, we identified the nearby protein-coding genes of the candidate SNP markers with the highest $-\log_{10}P$. We aligned the genes against the NCBI nr database by Blastx (Pruitt and Tatusova, 2005)

Results and Analysis

Identification of candidate loci by GWAS

The GWAS results from Group A were used to find the candidate genes. Although no genome-wide significant SNP for two traits was found according to the GWAS results in this study, a high proportion of genetic variance was explained by some regions. For example, some SNPs

with -$\log_{10}$P > 3 were concentrated in chromosome 4 and 11 for HUFA content. Five nearby SNPs with relative high -$\log_{10}$P in chromosome 4 explained 34.3% of the total genetic variance. The seven SNPs with -$\log_{10}$P > 3 were in a region spanning from 1.66 to 2.77 Mb of chromosome 11 for the trait of HUFA content. These seven SNPs explained 18.3% of genetic variance. Three SNPs with -$\log_{10}$P > 3 were concentrated in the region around 9.5 Mb in chromosome 11 and explained 28.2% of genetic variance. Details of these SNPs for Group A are presented in Table 4. These 15 SNPs were divided to three regions which were taken as the candidate regions to find candidate genes for n-3 HUFA content. As a results, *ATP8b1*, *FAF2*, *CERS2*, *FABP* and *DGAT1* genes were found in the three regions (shown in Table 4).

In the results of EW, one SNP named LG15_5148079 in chromosome 15 explained 4.02% of total additive genetic variance. Three SNPs with -$\log_{10}$P ≥ 4 on contig 000000268 from 187.4-187.5 kb region explained 17.3% of genetic variance. Three SNP loci concentrated in the region around 14 Mb in chromosome 6 explained 10.74% of genetic variance. All the regions and SNP mentioned above were taken as the candidate regions to find genes for EW trait, and the genes *HMX3*, *BORA*, *GRTP and IGF2* were found associated with the trait (shown in Table 5).

Association analysis of different groups in comparison

The results form *n-3* HUFA content that were analysed by GWAS using extreme phenotypes and all phenotypes have been compared from Group A to Group E. Table 2 shows the SNP loci and -$\log_{10}$P from highest to lowest in each group by GWAS. We used the 10 loci from Group A as a standard for assessing the differences from Group B to Group E. There were 6, 4, 5 and 2 identical loci with group A for groups B, C, D, and E, respectively. That is, 60%, 40%, 50%, and 20% of the SNP loci in Group A by using all the samples were detected by using 80%, 60%, 40% and 20% of the extreme phenotypes, respectively.

The SNP loci and the results of the significance tests for the trait of EW are presented in Table 3. According to the comparative analysis of the 10 loci in each group, we found that there were 4 identical loci between group E and group A, 4 between group D and group A, 6 between group C and group A, and 9 between group B and group A. In Group E, D, C, and B, 20%, 40%, 60%, and 80% of the extreme phenotypes were used to detect 40%, 40%, 60% and 90% of the candidate loci in Group A, respectively. The candidate loci did not increase in proportion to the numbers of extreme phenotypes or abide by a simple linear relationship. The adaptive 60% of the

extreme phenotypic subsample (group C) from the population may provide a method to save sequencing expenses and promote cost-saving studies.

Figure 1 shows the profiles of the Manhattan plot of GWAS for the trait of n-3 HUFA content in chromosome 1 to 24. We can find that the three candidate regions in Group A is also obvious in Group D, but is not obvious in Group E. In total, 3 candidate regions are identical on chromosome 11 for Groups A, B, C and D. Briefly, the suitable numbers of extreme phenotypes can locate the potential regions associated with phenotypes on the chromosome, such as the numbers of extreme phenotypes used in Group C or Group D. Figure 2 shows the results of GWAS in chromosome 1 to 24 for the trait of EW. The SNPs from the 10 significant loci of each group were distributed evenly on the chromosome. The distributions of the candidate regions were quite similar among Groups A, B and C.

Fig. 1 Manhattan plot of $-\log_{10}P$ for the genome-wide association study for the trait of HUFA content (Group A to Group E). X-axis represents the chromosome numbers (1–24). The last part is not the chromosome but represents SNPs have not been located on specific region in the genome.

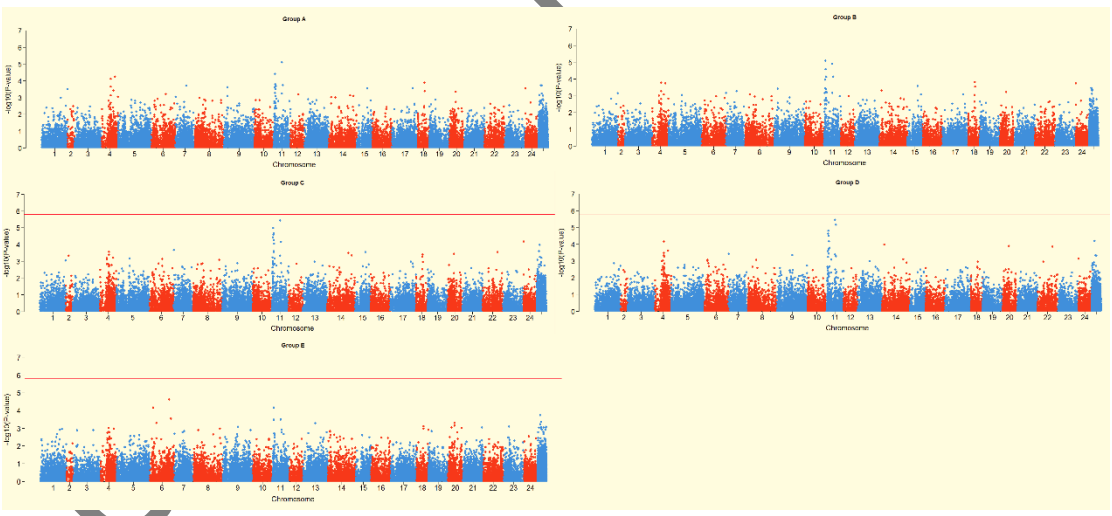
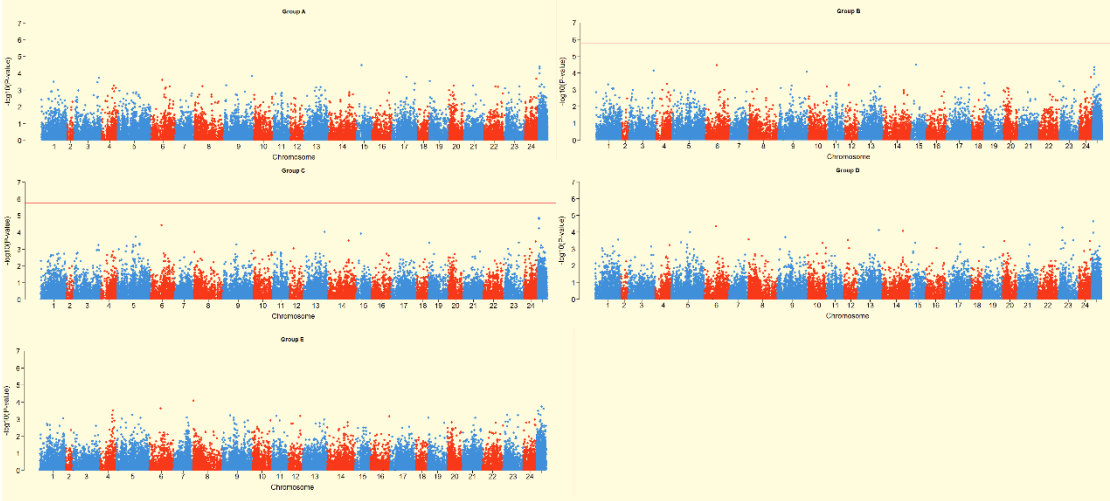


Fig. 2 Manhattan plot of $-\log_{10}P$ for the genome-wide association study for the trait of EW (Group A to Group E). X-axis represents the chromosome numbers (1–24). The last part is not the chromosome but represents SNPs have not been located on specific region in the genome.



Discussion

Both the economic traits of *n-3* HUFA content and EW are rarely directly measured from the live body. It is necessary to perform marker assisted selection for the two economic traits in the large yellow croaker. Thus, we performed the associated analysis for HUFA content and EW traits using SNPs information. In the present study, each trait was found 3 candidate regions (shown in Table 4 and 5) which were used to find candidate genes affecting the traits of HUFA and EW.

In particular, for the trait of HUFA, our results of extreme phenotypes confirmed the findings of previous studies, where one candidate region associated with HUFA content was located adjacent to the *FABP* gene (Fatty acid-binding protein), a generally accepted functional causal gene affecting fatty acid transport and fat accumulation (Darias *et al.* 2012; Duttaroy 2009; Guo *et al.* 2016). Another candidate region was relevant to the *DGAT* gene (Diacylglycerol Acyltransferase), a widely regulated gene with lipogenic enzyme activities, which are functional genes involved in the fat digestion and absorption pathway (Blanchard *et al.* 2016; Yamamoto *et al.* 2011). *CERS2* (ceramide synthase), which can regulate sphingolipid metabolism and fatty acid synthesis (Bauer *et al.* 2009;Brondolin *et al*, 2013; Mullen *et al.* 2012), which are the basis or pre-conditions of HUFA biosynthesis by the desaturase enzyme (Mohd-Yusof *et al.* 2010; Ren *et al.* 2013). *ATP8B1* gene has a relationship with candidate region on chromosome 4 and plays a role in lipid transport or metabolism (Haas *et al.* 2014). When the candidate regions were confirmed, we identified several novel candidate SNPs for the HUFA content trait from other groups as the supplement to detect functional genes (Table 6). Thus, extreme phenotypic sampling provides a supplement for the application of molecular breeding (marker-assisted selection)

through the identification of candidate region associated with functional genes. This was the first time that GWAS was performed for the EW trait to identify candidate regions and seek the potential functional genes in the large yellow croaker. Due to the limitation of sample size, the results may be identified as the first step in revealing the genetic background of the EW trait. The first candidate region was located close to the *HMX3* (Homeobox protein), which belongs to *HOX* gene family, which encode for transcription factors that regulate cell division (Zou and Jiang 2008). *HOX* genes specify cell fates in animal embryos, influence body weight gain (Amores and Postlethwait 1998; Lee *et al*. 2014) and are highly expressed in the tightly associated stromal connective tissues, as well as regional expression in tendons and muscle connective tissue. The research reveals a conserved role of phosphorylation of the N terminus of *BORA* for Plk1 activation and mitotic entry, which potential effected body growth as component of visceral weight for the large yellow croaker (Thomas *et al*. 2016). The second candidate region was located close to the gene *IGF2*, which effected body weight and muscle growth and was widely applied to fish breeding as an important candidate gene (Cleveland & Weber 2016; Kawanago *et al*. 2014). *GRTP1* (Growth hormone-regulated TBC protein) gene, the accurate function of which has not been detected. However, with a role as a growth hormone-regulated protein, it was valuable to study further to examine its effect on the phenotypes of EW trait of the large yellow croaker. The last candidate region was located on the *CYP1A1*. Maternal *CYP1A1* genotypes primarily effected on low birth weight in human when environment was polluted (Sram *et al*. 2006). Thus, we speculate *CYP1A1* genotypes effect large yellow croaker birth weight and body weight as the aquaculture environmental water deterioration. For the moment, it can be concluded that the EW trait is not driven by single gene with a major effect, but by various functional genes cooperating with each other.

Based on the results presented above, we performed various scales of extreme phenotypic sampling to detect the collection between the numbers and the result. For the economic trait of *n-3* HUFA content in this study, three candidate regions were identical between 76 extreme phenotypes and all 176 individuals (Group C and Group A). In other words, 60% of extreme phenotypic subsamples of the population detected a large proportion of candidate loci and located all of the main candidate genes, compared with using whole phenotypes. For the economic trait of EW, three candidate regions are common between the results analysed by 300 extreme phenotypes

and that by all phenotypes (Group C and Group A). The 300 extreme phenotypes (Group C) detect almost genome-wide SNPs loci associated with candidate genes, which is consistent with the results obtained using whole phenotypes (Group A). The results for 60% of extreme phenotypes subsample are similar with that of all phenotypes and as a reference to supplement the functional genes. However, the results did not increase simply in proportion to the numbers of extreme phenotypes. We speculate that use of approximately 60% of the extreme phenotypes subsample is capable of obtaining similar results at low cost-efficiency with GWAS by whole phenotypes of the large yellow croaker in the random population. The EP method ascertains the potential candidate region associated with economic traits, and extreme phenotypic sampling could save 40% of the cost for genotyping and DNA extraction, even though, the cost of sampling and trait measurement would not be economized.

Conclusions

This study compared the different numbers of extreme phenotypes for GWAS. The extreme phenotypic sampling found the potential association between SNP loci and economic traits. Each important economic traits respectively was detected three significant regions. We speculate 40%-60% of extreme phenotypic subsamples from the population could obtain a similar result as GWAS with whole phenotypes for further study. Thus, extreme phenotypic sampling could save 40% of the cost for genotyping and DNA extraction, without loss of the significant regions and candidate genes. Our study may provide a reference for others who want to perform GWAS with extreme phenotypes.

Ethics statement

The sample collection and experiments in the study was approved by the Animal Care and Use Committee of Fisheries College of Jimei University (Animal Ethics no. 1067).

Availability of data

Raw DNA sequencing reads were deposited in NCBI with the project accession of PRJNA309464 and SRA accession of SRR3114179.

Authors' contributions

XW conceived the research. ZW revised the manuscript. LW drafted the manuscript. LD performed the analysis and co-authored the manuscript. SX screened SNP loci. ZH analysed gene

function. Two reviewers revised our manuscript and gave very valuable advice. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank Kun Ye, Shuangbin Xu, Yuxue Gao and other colleagues in the laboratory that participated in fish sampling and measuring the traits. This work was supported by the Key Projects of the Xiamen Southern Ocean Research Centre (14GZY70NF34) and China Agriculture Research System (CARS-47-G04)

Reference

2016. *China Fishery Statistical Yearbook* **2016**: China Agricultural Press.

Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *NUCLEIC ACIDS RES.* **25**, 3389–3402

Amores A. and Postlethwait J. H. 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science*. **282**, 1711-1714.

Barnett I. J., Lee, S. and Lin, X. 2013. Detecting Rare Variant Effects Using Extreme Phenotype Sampling in Sequencing Association Studies. *Genetic Epidemiology*. **37,** 142–151.

Bauer R., Voelzmann A., Breiden B., Schepers U., Farwanah H., Hahn I. et al. 2009. Schlank, a member of the ceramide synthase family controls growth and body fat in Drosophila. *EMBO J*. **28**, 3706-3716.

Benjamini Y. and Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B.* **57***:*289–300

Blanchard P. G., Turcotte V., Côté M., Gélinas Y., Nilsson S., Olivecrona G. et al. 2016. Peroxisome proliferator-activated receptor γ a ctivation favours selective subcutaneous lipid deposition by coordinately regulating lipoprotein lipase modulators, fatty acid transporters and lipogenic enzymes. *Acta Physiol*. **217**, 227-239

Brondolin M., Berger S., Reinke M., Tanaka H., Ohshima T., Fuβ B. et al. 2013. Identification and Expression Analysis of the Zebrafish Homologs of the ceramide synthase Gene Family. *Dev Dyn*. **242**, 189–200.

Browning B. L. and Browning S. R. 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Am J Hum Genet*. **84**, 210-223.

Chen C., Ai H., Ren J., Li W., Li P., Qiao R. et al. 2011. A global view of porcine transcriptome in three tissues from a full-sib pair with extreme phenotypes in growth and fat deposition by paired-end RNA sequencing. *Bmc Genomics*. **12**, 448-448.

Cleveland B. M. and Weber G. M. 2016. Effects of steroid treatment on growth, nutrient partitioning, and expression of genes related to growth and nutrient metabolism in adult triploid rainbow trout (Oncorhynchus mykiss). *Domest Anim Endocrinol*. **56**, 1-12.

Darias M. J. and Boglino A. 2012. Molecular regulation of both dietary vitamin A and fatty acid

absorption and metabolism associated with larval morphogenesis of Senegalese sole (Solea senegalensis). *Comp Biochem Physiol A Mol Integr Physiol*. **161**, 130–139.

Dong L., Xiao S., Chen J., Wan L. and Wang, Z. 2016. Genomic Selection Using Extreme Phenotypes and Pre-Selection of SNPs in Large Yellow Croaker (Larimichthys crocea). *Mar Biotechnol*. **18**, 575-583.

Dong L., Xiao S., Wang Q. and Wang Z. 2016. Comparative analysis of the GBLUP, emBayesB, and GWAS algorithms to predict genetic values in large yellow croaker (Larimichthys crocea). *BMC Genomics*. **17**, 1.

Duttaroy A. 2009. Transport of fatty acids across the human placenta: a review. *Prog Lipid Res*. **48**, 52–61.

Feng Y, Xu Q. 2010. Pivotal role of hmx2, and hmx3, in zebrafish inner ear and lateral line development. *DEV BIOL*. **339**, 507.

Folch J., Lees M., Sloane-Stanley G. 1957. A simple method for the isolation and purification of total lipids from animal tissues. *J Biol Chem.* **226**, 497–509.

Guo H., Wei G., Peng S. et al. 2016, Transcriptome Analyses Reveal Lipid Metabolic Process in Liver Related to the Difference of Carcass Fat Content in Rainbow Trout (Oncorhynchus mykiss). *Int J Genomics*. **2016**, 1-10.

Haas M. J.,Shah G. N., Onsteadhaas L. M, et al. 2014. Identification of ATP8B1 as a blood-brain barrier-enriched protein. *CELL MOL NEUROBIOL*. **34**, 473-478.

Kang H. M., Sul J. H., Service S. K., et al. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet*. **42**, 348.

Kang H. M, Zaitlen N. A., Wade C. M., et al. 2008. Efficient control of population structure in model organism association mapping. *GENETICS*. **178**, 1709-1723.

Kawanago M., Takemura S., Ishizuka R., Kousaka T. and Shioya I. 2014. Leucine Affects Growth and Hepatic Growth-Related Factor Gene Expression in Japanese Amberjack. *N Am J. Aquacult*. **76**, 415-422.

Lee H. M., Rim H. K., Seo J. H., Kook Y. B., Kim S. K., Oh C. H. et al. 2014. HOX-7 suppresses body weight gain and adipogenesis-related gene expression in high-fat-diet-induced obese mice. *BMC Complement Altern Med*. **14**, 1-7.

Li D., Lewinger J. P., Gauderman W. J., Murcray C. E. and Conti D. 2011. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet Epidemiol*. **35**, 790–799.

Li H. and Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754-1760.

Mohd-Yusof N. Y., Monroig O., Mohd-Adnan A., Wan K. L. and Tocher D. R. 2010. Investigation of highly unsaturated fatty acid metabolism in the Asian sea bass, Lates calcarifer. *Fish Physiol Biochem*. **36**, 827-843.

Mullen T. D., Hannun Y. A. and Obeid, L. M. 2012. Ceramide synthases at the centre of sphingolipid metabolism and biology. *Biochem. J*. **441**, 789-802.

Murillo E., Rao K. S. and Durant A. A. 2013. The lipid content and fatty acid composition of four eastern central Pacific native fish species. *J. Food Compos Anal*. **33**, 1-5.

Omar D. L. C., Wen X., Ke B., Song M. and Nicolae D. L. 2009. Gene, region and pathway level analyses in whole-genome studies. *Genet Epidemiol.* **34**, 222-231.

Pruitt K., Tatusova T. D. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence

database of genomes, transcripts and proteins. *NUCLEIC ACIDS RES*, **33**: 501-504.

Ren H. T., Zhang G. Q., Li J. L., Tang Y. K., Li H. X., Yu J. H. et al. 2013. Two Δ6-desaturase-like genes in common carp (Cyprinus carpio var. Jian): Structure characterization, mRNA expression, temperature and nutritional regulation. *Gene*. **525**, 11-17.

Segura V., Vilhjálmsson B. J., Platt A., Korte A., Seren Ü., Long Q. et al. 2012. An efficient multi-locus mixed model approach for genome-wide association studies in structured populations. *Nat Genet*. **44**, 825-830.

Shi J. and Walker M. G. 2007. Gene set enrichment analysis (GSEA) for interpreting gene expression profiles. *CURR BIOINFORM*. **2**, 133–137

Sram R. J., Binkova B., Dejmek J. et al. 2006. Association of DNA adducts and genotypes with birth weight. *MUTAT RES-GEN TOX EN*, **608**, 121-128.

Thomas Y., Cirillo L., Panbianco C. et al. 2016. Cdk1 Phosphorylates SPAT-1/Bora to Promote Plk1 Activation in C. elegans and Human Cells. *CELL REP*. **15**, 510.

Xiao S., Wang P., Dong L., Zhang Y., Han Z., Wang Q. and Wang, Z. 2016. Whole-genome single-nucleotide polymorphism (SNP) marker discovery and association analysis with the eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) content by Genotyping-By-Sequencing (GBS) in teleost Larimichthys crocea. *Peerj*. **4**

Yamamoto T., Yamaguchi H., Miki H., Kitamura S., Nakada Y., Aicher T. D. et al. 2011. A novel coenzyme A:diacylglycerol acyltransferase 1 inhibitor stimulates lipid metabolism in muscle and lowers weight in animal models of obesity. *Eur J. Pharmacol*. **650**, 663-672.

Yang J., Jiang H., Yeh C. T., Yu J., Jeddeloh J. A., Dan N. et al. 2015. Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. *Plant. J*. **84**, 222-226.

Zhang X., Huang S., Zou F. and Wang W. 2010. TEAM: efficient two-locus epistasis tests in human genome-wide association study. *Bioinformatics*. **26**, 217–227.

Zhou Y. J., Wang Y. and Chen L. L. 2016. Detecting the Common and Individual Effects of Rare Variants on Quantitative Traits by Using Extreme Phenotype Sampling. *Genes*. **7**, 1-12.

Zou S. and Jiang X. 2008. Gene duplication and functional evolution of Hox genes in fishes. *J. Fish Biol*. **73**, 329-354.

Table 1. Experimental design of phenotypic data for the traits of HUFA and EW.

Group	Numbers	
	HUFA	**EW**
A group (100%)[b]	176 [a]	500
B group (80%)	140	400
C group (60%)	106	300
D group (40%)	70	200
E group (20%)	36	100

[a] The extreme phenotypes contain a half numbers of the highest values and a half numbers of the lowest values.

[b] The Proportion of extreme phenotypes from the samples.

Table 2. Ten candidate SNP loci of HUFA in different groups

SNP marker	A group--log10P	SNP marker	B group--log10P	SNP marker	C group--log10P	SNP marker	D group--log10P	SNP marker	E group--log10P
LG11_907670 4_snp[a]	5.1196761	LG11_183004 0_snp	5.0934653	LG11_907670 4_snp	5.1473529	LG11_907670 4_snp	5.4640832	LG6_1920095 1_snp	4.6344371
LG11_183004 0_snp	4.409677	LG11_907670 4_snp	4.9178419	LG11_183004 0_snp	4.9850912	LG11_992391 0_snp	5.1637435	LG6_2976308 _snp	4.1668334
LG4_1309496 0_snp	4.2540693	LG11_225106 2_snp	4.6001472	LG11_256653 3_snp	4.6664748	LG11_225106 2_snp	4.8035464	LG6_2976303 _indel	4.1668334
LG4_8947695 _snp	4.1296821	LG11_256653 3_snp	4.1412401	LG11_225106 2_snp	4.590117	LG11_183004 0_snp	4.6253886	LG6_2976304 _indel	4.1668334
LG18_676173 6_snp	3.8917465	LG11_992391 0_snp	4.1370468	LG11_165814 9_snp	4.4372506	LG11_256653 3_snp	4.5060382	LG6_2976306 _snp	4.1668334
LG11_225106 2_snp	3.8073673	LG11_276694 4_snp	4.1325326	LG11_256652 2_snp	4.275075	000000611_3 31851_snp	4.1941385	LG11_183004 0_snp	4.1566787
LG11_992391 0_snp	3.7570024	LG11_165814 9_snp	3.9585176	LG24_233618 _snp	4.1823721	LG11_276694 4_snp	4.1794062	000000611_3 31851_snp	3.7520673
000000688_5 04816_snp	3.7390454	LG18_676173 6_snp	3.821496	LG11_992391 0_snp	4.1671925	LG4_8947695 _snp	4.1656194	LG6_2096648 9_snp	3.5520124
000000475_1 47190_snp	3.7365925	LG4_8947695 _snp	3.7947509	LG11_276694 4_snp	4.056695	LG14_223234 3_snp	3.9832821	LG11_907670 4_snp	3.4910684
LG7_1106081 2_snp	3.7064962	LG24_233618 _snp	3.7558491	000000611_3 31851_snp	3.9816835	LG20_659791 3_snp	3.8880646	LG11_188534 8_indel	3.4761183

[a] The tag of the candidate Genome-wide SNPs for the trait of HUFA with different colours

Table 3. Ten candidate SNP loci of EW in different groups

SNP marker	A group--log10P	SNP marker	B group--log10P	SNP marker	C group--log10P	SNP marker	D group--log10P	SNP marker	E group--log10P
LG15_5148079_snp	4.4951	LG15_5148079_snp	4.5067	000000268_187482_snp	4.8729	000000268_187482_snp	4.6410	LG8_554597_snp	4.0854
000000268_187482_snp	4.3887	LG6_11321041_snp	4.4616	000000268_187467_snp	4.8034	000000268_187467_snp	4.6410	000002898_96026_snp	3.7370
000000268_187467_snp	4.2894	000000268_187482_snp	4.3374	LG6_11321041_snp	4.4366	LG6_11321041_snp	4.3697	LG6_11321041_snp	3.6228
000000268_187460_snp	4.0040	000000268_187467_snp	4.1799	000000268_187460_snp	4.2449	LG23_4551748_snp	4.2735	000003175_566889_snp	3.6028
LG9_29194612_snp	3.8354	LG3_25452265_snp	4.1447	LG13_21464486_snp	4.0417	LG13_21464486_snp	4.1216	LG4_12884066_snp	3.4947
LG17_14982207_snp	3.7922	LG9_29194612_snp	4.0734	LG15_5148079_snp	3.9365	LG14_20924173_snp	4.0743	000000268_187482_snp	3.4744
LG3_25452265_snp	3.7275	000000268_187460_snp	3.9543	LG5_18734236_snp	3.7460	LG14_20924169_indel	4.0743	000000268_187467_snp	3.4744
LG24_11624327_snp	3.6709	LG24_11624327_snp	3.7571	LG14_20924169_indel	3.5196	LG5_18734236_snp	3.9977	000000268_187460_snp	3.4744
LG6_11321041_snp	3.6140	LG23_671133_snp	3.4994	LG14_20924173_snp	3.5196	000000268_187460_snp	3.9596	000000364_863660_snp	3.3228
LG19_556547_snp	3.5405	LG19_556527_snp	3.3815	LG24_11624327_snp	3.4640	LG9_8088607_indel	3.6936	LG5_16442582_snp	3.2668

[a]The tag of the candidate Genome-wide SNPs for EW trait with different colours

441
442
443

Table 4. Genome-wide candidate SNPs for HUFA trait

SNP	Chromosome	Position[a](bp)	Nearby gene[b]	Distance (bp)	-log10P [c]	Function
LG11_9076704_snp	11	9076704	FABP	288290	5.12	Lipid transport and metabolism
LG11_9076721_snp	11	9076721	FABP	288273	3.24	Lipid transport and metabolism
LG11_9923910_snp	11	9923910	FABP	558916	3.76	Lipid transport and metabolism
LG11_1658149_snp	11	1658849	DGAT 1	598693	3.60	Lipid transport and metabolism
LG11_1830040_snp	11	1830040	DGAT 1	426802	4.41	Lipid transport and metabolism
LG11_1885348_snp	11	1885348	DGAT 1	371494	3.29	Lipid transport and metabolism
LG11_2251062_snp	11	2251062	DGAT 1	5780	3.81	Lipid transport and metabolism
LG11_2566533_snp	11	2566533	DGAT 1	309691	3.51	Lipid transport and metabolism
LG11_2638790_snp	11	2638790	DGAT 1	381948	3.09	Lipid transport and metabolism
LG11_2766944_snp	11	2766944	DGAT 1	510102	3.70	Lipid transport and metabolism
LG4_8782255_snp	4	8782255	ATP8B1	340781	3.66	Phospholipid transporting
LG4_8947695_snp	4	8947695	ATP8B1	175341	4.13	Phospholipid transporting
LG4_8928457_snp	4	8928457	ATP8B1	194579	3.14	Phospholipid transporting

LG4_11630539_snp	4	11630539	FAF2	238456	3.43	Lipid transport and metabolism
LG4 13094960_snp	4	13094960	CERS2	176712	4.25	Regulative lipidic metabolism

[a] The position on the large yellow croaker genome sequence.

[b] The nearby gene to the SNP.

[c] -log10 (p-value) from Group A

Table 5. Genome-wide significant SNPs for eviscerated weight (EW) trait

SNP	Chromosome	Position[a](bp)	Gene[b] Name	Distance[a](bp)	$-\log_{10}P$	Function
LG15_5148079	15	5148079	HMX3	193878	4.50	Growth
LG15_5148079	15	5148079	EGR4	62945	4.50	Growth
000000268_187482_snp	mismatching	187482	CYP1A1	137580	4.39	Birth weight
000000268_187467_snp	mismatching	187467	CYP1A1	137595	4.29	Birth weight
000000268_187460_snp	mismatching	187460	CYP1A1	137602	4.00	Birth weight
LG6_11321041	6	11321041	IGF 2	3610596	3.61	Growth
LG6 15112475 C	6	15112475	IGF 2	180838	3.15	Growth
LG6 15112475 C	6	15112475	GRTP1	10626	3.15	Growth
LG6_14388415_snp	6	14388415	IGF 2	904898	3.08	Growth

[a] The position on the large yellow croaker genome sequence.

[b] The nearby known gene to SNP.

463

464 **Table 6. Genome-wide significant SNPs for HUFA trait in different groups**

SNP	Chromosome	Position[a](bp)	Nearby gene[b]	Distance (bp)	-log10P [c]	Function
LG11_2566522_snp	11	2566522	DGAT 1	309680	4.27 (Groupo C)	Lipid transport and metabolism
LG11_2566605_snp	11	2566605	DGAT 1	309763	3.59 (Groupo C)	Lipid transport and metabolism
LG11 3370511_snp	11	3370511	CERS2	With in	3.78 (Groupo D)	Regulative lipidic metabolism
LG14_2232343_snp	14	2232343	DGAT 2	199765	3.98 (Groupo D)	Lipid transport and metabolism
LG9_15698414_snp	9	15698414	elov15	95370	3.09 (Groupo E)	Lipid transport and metabolism

465 [a] The position on the large yellow croaker genome sequence.

466 [b] The nearest known gene to the significant SNP.

467 [c] -log10 (p-value) from Group

468

469