Pak. J. Agri. Sci., Vol. 54(1), 201-208; 2017
ISSN (Print) 0552-9034, ISSN (Online) 2076-0906
DOI: 10.21162/PAKJAS/17.3602
http://www.pakjas.com.pk

SINGLE NUCLEOTIDE POLYMORPHISMS IDENTIFIED FROM BLUNT SNOUT BREAM (*Megalobrama amblycephala*) TRANSCRIPTOME THROUGH NEXT GENERATION SEQUENCING

Ngoc Tuan Tran[1,3], Ze-Xia Gao[1,4], Hong-Hao Zhao[1,4] and Wei-Min Wang[1,2,*]

[1]College of Fisheries, Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education/Key Lab of Freshwater Animal Breeding, Ministry of Agriculture, Huazhong Agricultural University, Wuhan, Hubei 430070, China; [2]Collaborative Innovation Center for Efficient and Health Production of Fisheries in Hunan Province, Changde 41500, China; [3]Institute of Hydrobiology, Chinese Academy of Sciences, and the Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Wuhan, Hubei Province, P.R. China; [4]Freshwater Aquaculture Collaborative Innovation Center of Hubei Province, Wuhan 430070, China.
[*]Corresponding author's e-mail: wangwm@mail.hzau.edu.cn

Single nucleotide polymorphisms (SNPs) are the most dominant type of DNA variations in genome which is useful in genetic linkage mapping and quantitative traits loci analysis. In this study, a SNP's database of blunt snout bream (*Megalobrama amblycephala*) was obtained from the transcripts of cDNA library using Solexa/Illumina sequencing. A total of 36,326 putative SNPs were discovered from the transcripts. One SNP was found in 29.2 base pair length of the transcripts. The transition and transversion mutation was 21,445 and 12,553 SNPs, respectively. A ratio of transition to transversion was 1.71. Within those yielded SNPs, 10,812 SNPs identified from 2,421 unigenes could be annotated to differential functionality in comparing to public database using BLASTX for gene ontology (GO), eukaryotic orthologous groups of proteins (KOG), and Kyoto encyclopedia of genes and genomes (KEGG). A number of SNPs (n = 7,727, 71.5%) found from 1,628 unigenes were assigned to three main GO categories: 'cellular components', 'molecular function' and 'biology process'. In total, 5,812 SNPs (53.8%) identified from 1,324 unigenes were classified into 25 KOG categories. 4,589 SNPs (42.4%) detected from 975 unigenes were assigned to 278 KEGG pathways. Furthermore, a number of 600 SNPs found from 111 unigenes were successfully annotated in the term 'immune system' *via* KEGG classification. The SNP's database in this study could be useful for further genetic studies in blunt snout bream.

Keywords: *Megalobrama amblycephala*, SNP, transcriptome, next generation sequencing

INTRODUCTION

The polymorphic genetic markers are important in genetic-related researches such as population genetic structuring, relatedness, the genetic basis of adaptive traits, and genetic selection (Beaumont, 2005; Luikart *et al.*, 2003). Single nucleotide polymorphisms (SNPs) are the most dominant type of DNA variations in genome which is useful in animal genetic studies (Helyar *et al.*, 2011; Hinds *et al.*, 2005; Vignal *et al.*, 2002; Yu *et al.*, 2014). The genetic polymorphisms located in innate immune genes have been reported crucially in the functions of resistance or susceptibility in the host's immunity stimulated by pathogen infections (de Boer *et al.*, 2011), particularly, in the inflammation mechanisms (Hartel *et al.*, 2004). In aquaculture, SNPs found in coding regions of genes were useful for understanding the immune responses against pathogens and other stressful stimuli (Núñez-Acuña and Gallardo-Escárate, 2013). Moreover, SNP markers are previously considered to be important for genetic linkage mapping and quantitative trait loci analysis (He *et al.*, 2003; Kongchum *et al.*, 2010; Kucuktas *et al.*, 2009; Salem *et al.*, 2012; Wang *et al.*, 2012a; Wang *et al.*, 2012b). Thus, studies on SNPs in genes involving in differential functionality in several aquatic species have been conducted (Helyar *et al.*, 2011; Kongchum *et al.*, 2010; Núñez-Acuña and Gallardo-Escárate, 2013; Yu *et al.*, 2014).

Blunt snout bream *Megalobrama amblycephala* is a favorable freshwater aquaculture species in China because of high delicacy, high larval survival rate in breeding, natural foods feeding habitat, fast growth, and tender flesh (Tsao, 1960; Zhou *et al.*, 2008). In recent years, the production of this fish species has rapidly increased and attained the seventh among many Chinese freshwater fish species in China (MAPRC, 2012). Many breeding and selection programs of blunt snout bream which were mainly focused on its growth characteristics have been developed in China (Li and Cai, 2003; SiFa and WanQi, 2000). In farming practices, the most significant factor affecting farmed blunt snout bream is bacterial infection which caused high mortality during culture period inducing great economic losses (Ming *et al.*, 2012; Nielsen *et al.*, 2001), for example, the infection of *Aeromonas hydrophila* causing hemorrhagic septicemia in fish body and

changes in histopathological characteristics (He *et al.*, 2006; Nielsen *et al.*, 2001; Tran *et al.*, 2015a). Although, analysis of transcriptomic profile and identification of markers (microsatellite and SNPs) associated with growth traits of blunt snout bream has been previously implemented (Gao *et al.*, 2012), regarding SNP markers related to immunity in this fish species was insufficiently understood.

In this study, SNPs were identified from the transcriptomic database of blunt snout bream (Tran *et al.*, 2015b), which was obtained using Solexa/Illumina sequencing method, and then the differential functionality of these SNPs was characterized. The results of this study may be useful for further studies on development and applications of SNP markers in breeding and selection of blunt snout bream.

MATERIALS AND METHODS

***Fish and sample collection*:** Eighteen healthy blunt snout bream, with an average body weight of 41.5±15.6 g, were collected from a culture farm in Tuanfeng, Huanggang, Hubei Province, China. The fish were acclimatized for two weeks in the laboratory at the College of Fisheries, Huazhong Agricultural University, at a water temperature of about 28 °C. The fish were fed a commercial pelleted feed twice a day (8:00 and 16:00 hrs.). Before sampling, fish were anesthetized with tricaine methanesulfonate (MS-222, Sigma, USA) at 100 mg/L and surgically dissected. Six tissues (blood, liver, gill, intestine, spleen and kidney) were separately collected, immediately kept into liquid nitrogen, and stored at -80^{o}C until total RNA extraction.

***Total RNA extraction and Illumina sequencing*:** The total RNA of each sample was isolated with RNAiso Plus Reagent (Takara Bio Inc), according to the manufacturer's instructions. The quality of total RNA was evaluated using electrophoresis in 1% agarose gels. The quantity of total RNA was measured using a Nanodrop 2000 (Thermo Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies). Equal amounts of the total RNA from each of the six tissues were dissolved in RNase-free water and pooled in equal quantities to generate a pooled sample. The poly(A)$^{+}$ RNA was purified from the total cellular RNA using poly(dT) oligo-attached magnetic beads, and the full-length cDNAs were synthesized with the TruSeq RNA Sample Preparation Kit (Illumina Inc., USA), according to the manufacturer's protocol. The cDNA libraries were sequenced on the Illumina HiSeq2000 genomic sequencing platform to generate 100-bp paired-end reads, by Shanghai OE Biotech Company.

***SNP discovery*:** The reads from transcriptome were trimmed using sickle tool; then cleaned reads were assembled *de novo* using Trinity (Version r2013_08_14) (Grabherr *et al.*, 2011); and the assembled sequences was used as resources in this study. All assembled reads were aligned using SSAHA2 (Ning *et al.*, 2001) with default parameter and putative SNPs were identified using VarScan with default parameter. A mutation position with at least two SNP alleles was defined as a putative SNP or indels (insertion or deletion) after matching with four or more sequences. For high quality SNP identification, all unigenes containing SNPs were screened with the criteria such as read depth, and the quality of flanking regions and absence of other SNPs within 15 bp flanking regions (Gao *et al.*, 2012).

***Functional annotation of unigenes containing SNPs*:** The unigenes containing SNPs from the transcriptome profile were compared against the Swissprot database using BLASTX with the top hit of a sequence similarity >30% and an e-value cut-off of 1e-5. With BLASTX results, all obtained unigenes were assigned to gene onthology (GO) annotation with three main categories, described as 'biological processes', 'molecular functions' and 'cellular components' (http://www.geneontology.org/). The unigenes were aligned to the eukaryotic orthologous groups of proteins (KOG) database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to predict and identify the possible functional categories of the genes. Kyoto encyclopedia of genes and genomes (KEGG) pathway mapping (http://www.genome.jp/kegg/) was used to determine the biological pathways involved (Kanehisa *et al.*, 2004). The KEGG pathway was annotated using the KEGG automatic annotation server (KAAS) (http:/www.genome.jp/tools/kaas/) based on KEGG orthology (KO) identifiers (Moriya *et al.*, 2007).

***Identification of SNPs involving in 'immune system' pathways*:** The immune-related genes were discovered based on the Swissprot database by setting up the threshold of an e-value cut-off of 1e-30 (Núñez-Acuña and Gallardo-Escárate, 2013). The identification of SNPs locating in different immune-related genes in pathways was performed following assigned to the term 'immune system' from the KEGG database (Kanehisa *et al.*, 2004).

RESULTS

Transcriptome analysis and SNP identification: A total of 57,668,696 raw reads (total length of 5,766,869,600 base pair (bp)) of blunt snout bream cDNA library was gained using Solexa/Illumina sequencing method. After trimming and removing lower quality reads, a total of 48,046,044 reads were obtained from the database. The trimmed reads covered 4,616,027,243 bp, generating an average of 96.1 bp per a read. The transcriptome was assembled *de novo* with Trinity software (Grabherr *et al.*, 2011). *De novo* assembly produced a number of 7,862 unigenes (high quality checking of more than 200 bp in a sequence) containing polymorphic bases.

As a result, a total of 36,326 putative SNPs distributing in 7,862 unigenes were obtained from the blunt snout bream transcriptome profile. One SNP was predicted in a 29.2 bp length of the sequences. Within these detected SNPs, a higher amount of SNPs were found in transition (21,455 SNPs) than

such in transversions (12,553 SNPs). A ratio of transition over transversion was 1.71. The number of each mutation type in the two transition types (A/G and C/T) and the four transversion types (A/T, A/C, G/T, and C/G) was approximately similar, respectively (Fig. 1). Furthermore, the read depth in SNPs position was 11.8 in average (ranging from 2 to 100); in which, the read depth within 10 account was highest (including 24,560 SNPs, 67.6%), followed by between 10 and 50 account and more than 50 account was 10,555 (29.1%) and 1,211 SNPs (3.33%), respectively. PCR primers were successfully designed for 33,337 of the identified SNPs.

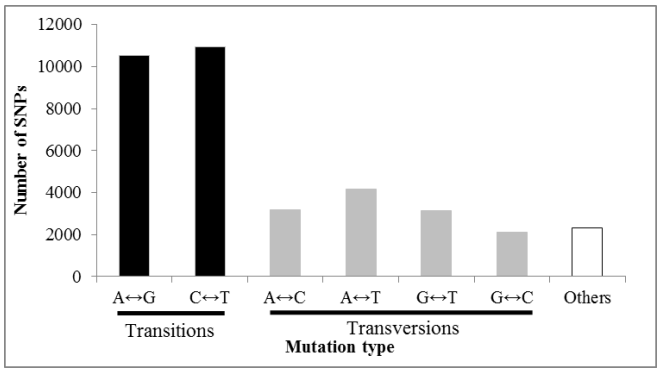

Figure 1. Distribution of different types of SNPs from the transcriptome profile.

***SNP functional annotation and analysis*:** The BLASTX was used to query the protein database at NCBI with the top hit of a sequence similarity >30% and an e-value cut-off of 1e-5. As a result, a total of 2,421 unigenes (30.8%) containing 10,812 SNPs in the database showed a significant hit and were used to annotate differential functionality.

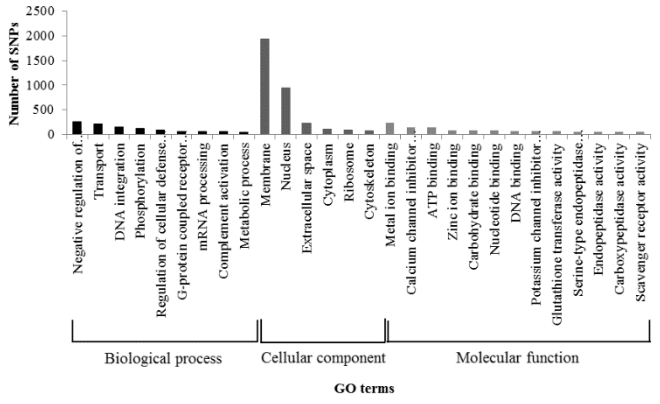

Figure 2. GO annotation of SNP number distributing in unigenes.

In GO annotation only 1,628 unigenes (67.2%) containing 7,727 SNPs (71.5%) were annotated and assigned to three main GO categories ('biology process', 'cellular components' and 'molecular function') including 301 terms (Table S1). The maximum number of SNPs (3,854, 49.9%) allocated in the 'cellular component' category. Following this category, 2,018 (26.1%) and 1,855 SNPs (24.0%) were found contributed to the 'molecular function' and 'biology process' category, respectively. The most 34 terms of GO classification comprising more than 50 polymorphisms in GO annotation were shown in Figure 2. Three terms of 'negative regulation of peptidase activity' (including 263 SNPs), 'transport' (215 SNPs) and 'DNA integration' (155 SNPs) were determined containing the most abundant SNP number in the 'biological process' category. Three terms of 'membrane' (1,935 SNPs), 'nucleus' (941 SNPs) and 'extracellular space' (226 SNPs) included the most common SNP number in the 'cellular component' category. Three terms of 'metal ion binding' (229 SNPs), 'calcium channel inhibitor activity' (145 SNPs) and 'ATP binding' (139 SNPs) comprised the most popular number of SNPs in the 'molecular function' category.

In KOG annotation, only 1,324 unigenes, including 5,812 SNPs (53.8%) were annotated by KOG database (Table S2). Among these achieved SNPs, a total of 5,143 SNPs locating on 1,181 unigenes were assigned to 24 KOG functional categories (Fig. 3). Of which functional categories, the category 'transcription' (coded 'K') contained the highest number of SNPs (627 SNPs, contributing 12.2%).

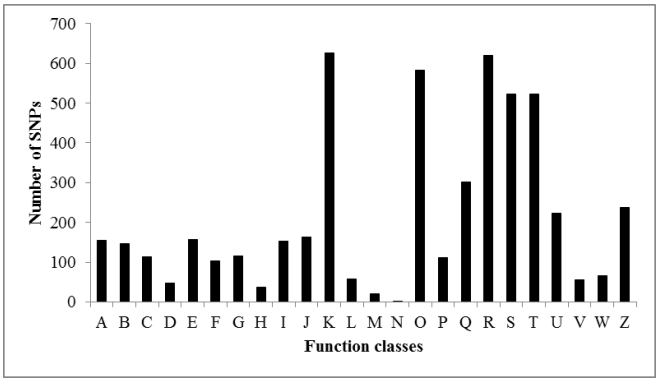

Figure 3. KOG annotation of number of SNPs presented in unigenes. [A] RNA processing and modification; [B] Chromatin structure and dynamics; [C] Energy production and conversion; [D] Cell cycle control, cell division, chromosome partitioning; [E] Amino acid transport and metabolism; [F] Nucleotide transport and metabolism; [G] Carbohydrate transport and metabolism; [H] Coenzyme transport and metabolism; [I] Lipid transport and metabolism; [J] Translation, ribosomal structure and biogenesis; [K] Transcription; [L] Replication, recombination and repair; [M] Cell wall/membrane/envelope biogenesis; [N] Cell motility; [O] Posttranslational modification, protein turnover, chaperones; [P] Inorganic ion transport and metabolism; [Q] Secondary metabolites biosynthesis, transport and catabolism; [R] General function prediction only; [S] Function unknown; [T] Signal transduction mechanisms; [U] Intracellular trafficking, secretion, and vesicular transport; [V] Defense mechanisms; [W] Extracellular structures; [Z] Cytoskeleton.

Table 1. Summary of number of SNPs located in unigenes by KEGG annotation.

KEGG categories	Terms	Number of SNPs
Human diseases	Infectious diseases	2609
	Cancers	1494
	Cardiovascular diseases	445
	Immune system diseases	434
	Neurodegenerative diseases	141
	Metabolic diseases	89
Organismal systems	Immune system	1224
	Nervous system	828
	Digestive system	651
	Endocrine system	297
	Development	145
	Environmental adaptation	129
	Excretory system	104
	Circulatory system	22
	Sensory system	22
Metabolism	Carbohydrate metabolism	893
	Lipid metabolism	861
	Xenobiotics biodegradation and metabolism	767
	Metabolism of cofactors and vitamins	468
	Glycan biosynthesis and metabolism	334
	Amino acid metabolism	220
	Nucleotide metabolism	214
	Energy metabolism	155
	Metabolism of other amino acids	134
	Biosynthesis of other secondary metabolites	13
	Metabolism of terpenoids and polyketides	13
Cellular processes	Transport and catabolism	519
	Cell communication	423
	Cell growth and death	263
	Cell motility	160
Environmental information processing	Signal transduction	1041
	Signaling molecules and interaction	283
	Membrane transport	44
Genetic information processing	Translation	494
	Folding, sorting and degradation	477
	Transcription	169
	Replication and repair	85

Other categories such as 'general function prediction only' (coded 'R') (620, 12.1%), 'posttranslational modification, protein turnover, chaperones' (coded 'O') (583, 11.3%), and 'signal transduction mechanisms' (coded 'T') (524, 10.2%) included a high number of SNPs. The category 'cell motility' (coded 'N') tended to be less SNP number (1, 0.02%). Besides, still 669 SNPs distributing in 142 unigenes were annotated to be at least two categories of a total of 25 KOG's functionality (Table S2).

In KEGG annotation, there were 975 unigenes containing 4,589 SNPs (42.4%) were classified into 278 KEGG pathways (Table S3). The pathways were assigned to six KEGG categories, comprising of 'human diseases', 'organismal systems', 'metabolism', 'cellular processes', 'environmental information processing', and 'genetic information processing' (Table 1). As showed in Table 1, the more number of SNPs were found in 'human diseases' KEGG category. Seventeen pathways with the highest number of SNPs (more than 200 SNPs locating in each pathway) were illustrated in Figure 4.

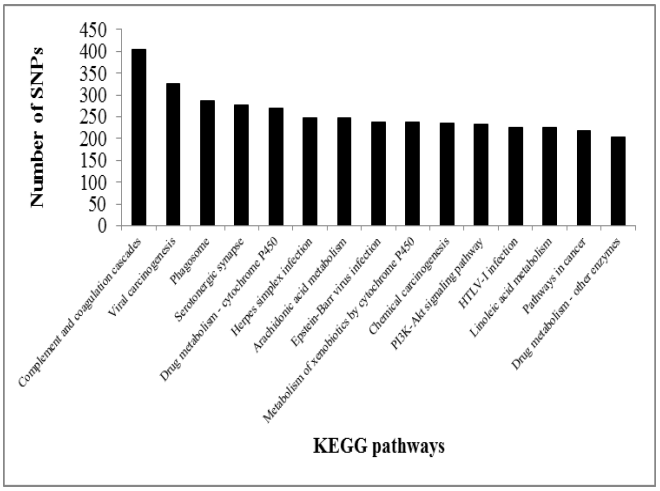

Figure 4. Distribution of SNPs obtained in 18 KEGG pathways.

***SNPs related to 'immune system' pathways*:** A number of 2,054 unigenes including 8,483 SNPs was yielded from the Swissprot database (threshold of an e-value cut-off of 1e-30). Among these unigenes, only unigenes encoded genes participating to pathways involving in the term 'immune system' in the KEGG database were selected to investigate. As a result, a number of 111 unigenes (5.4%) including 600 SNPs were successfully annotated in the term 'immune system'. The distribution of these SNPs in each pathway was illustrated in Figure 5. Of which yielded SNPs, 573 SNPs (95.5%) were found belonging to the transition and transversion types of mutation, shown in the Figure 6. The ratio of transition over transversion was 1.67, implying that each type of transitional mutation is generated 3.34 times as often as each type of transversional mutation (Collins and Jukes, 1994). The study resulted in alpha-2-macroglobulin (A2M) contained the highest number of SNPs (112 SNPs), followed by toll-like receptor 5 (TLR5) and complement

component 3 (C3), which had 33 and 32 SNPs, respectively. The remaining genes with a number of SNP contribution was shown in Table S4. Different KEGG pathways including the maximum number of SNPs located in different immune-related genes were used to describe specifically.

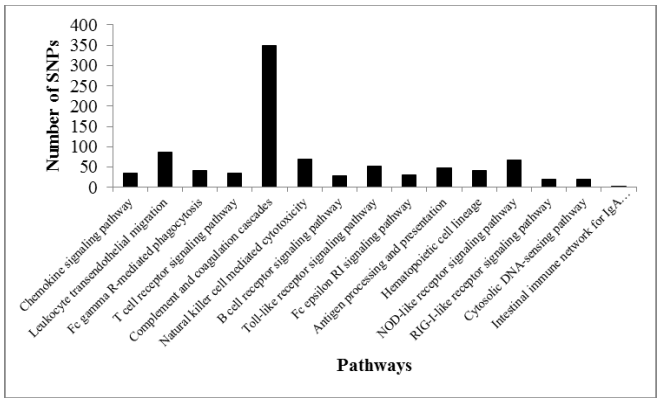

Figure 5. Distribution of SNP number in pathways in the term 'immune system'.

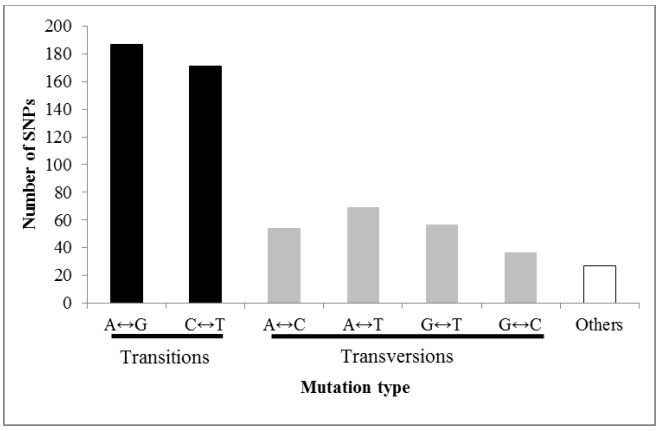

Figure 6. Distribution of different types of SNPs identified from unigenes assigned to the term 'immune system'.

In the complement and coagulation cascades pathways, there were 21 genes (encoded by 36 unigenes), including 350 SNPs were determined (Fig. 7A). Of these genes, A2M had the

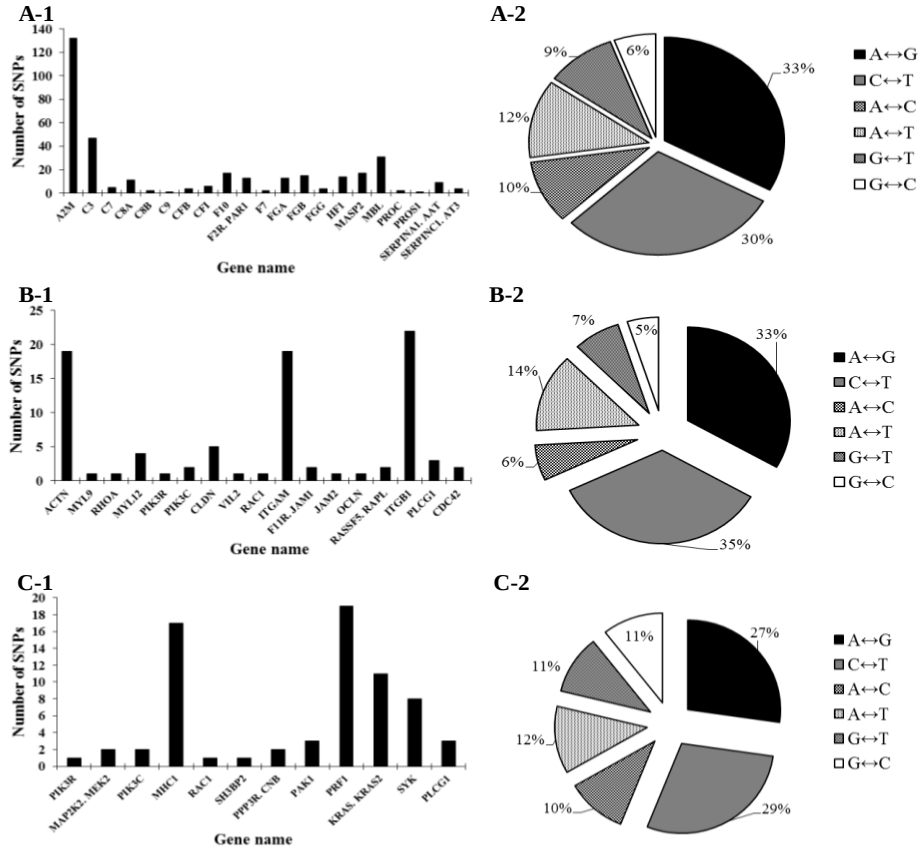

Figure 7. Description of SNPs presented in the complement and coagulation cascades pathway (A), leukocyte transendothelial migration pathway (B), and natural killer cell mediated cytotoxicity pathway (C). (*-1): Distribution of SNPs by different genes of the pathways and (*-2): Distribution of SNPs by kinds of polymorphism.

highest number of SNPs, containing 132 SNPs (37.7%). This gene was followed by C3 and mannose-binding lectin (MBL) which correspondingly comprised 47 (13.4%) and 31 SNPs (4.86%). The two mutation types of transition and transversion were common (96.3%), comparing with indels. Among these mutation types, the SNPs were clear bias towards the transition (A↔G: 32% and C↔T: 30%).

In the leukocyte transendothelial migration pathway, a total of 17 genes (mapped by 22 unigenes) with 87 SNPs located were recorded (Fig. 7B). The integrin beta 1 (ITGB1) was the most abundant distribution of SNPs with 25.3% (22 SNPs) of the whole number of SNPs distributed in this pathway. Following this gene, two genes actinin alpha (ACTN) and integrin alpha M (ITGAM) were found with a higher number of SNPs, including 19 SNPs for each. The variation mechanisms were more popular with transition and transversion (93.1%) than other types of mutation (indels). The more transition mutation type (68%) was presented in this pathway.

In the natural killer cell mediated cytotoxicity pathway, a total of 70 SNPs were gathered from 12 different genes encoded by 13 unigenes (Fig. 7C). The maximum number of SNPs obtained in this pathway was discovered in perforin 1 (PRF1) which had 19 SNPs (27.1%). It was followed by major histocompatibility complex class I (MHC I) (17 SNPs, 24.3%), GTPase KRas (KRAS.KRAS2) (11, 15.7%) and spleen tyrosine kinase (SYK) (8, 11.4%). In this pathway, the transition and transvertion mutation types were also dominant which occupied 94.3% compared with indels. The SNPs database gave a transition to transvertion ratio of 1.28.

DISCUSSION

This is the first study on identification of SNPs locating in immune-related genes using transcriptome analysis profile in blunt snout bream. In this study, the Solexa/Illumina sequencing reads obtained from cDNA library of blunt snout bream were used for identification of SNPs. Herein, a total of 36,326 putative SNPs were discovered from a number of 7,862 unigenes of the blunt snout bream transcriptomic profile. One SNP was found in a sequence length of 29.2 bp. This is much lower than previous finding of one SNP every 302 bp reported from transcriptomic database of blunt snout bream (*M. amblycephala*) using 454 pyrosequencing (Gao *et al.*, 2012). The ratio of transition over transversion plays an important role in genes identification after affected by generation selection (Jung *et al.*, 2011; Morton *et al.*, 2006). Interestingly, the ratio of transition over transversion (1.71) in this study is accordance with the previous report on such ratio calculated in database of human SNPs identified from EST sequences (Picoult-Newberg *et al.*, 1999). However, the transition over transversion ratio in current study is lower than which those in *Macrobrachium rosenbergii* transcriptome profile (the ratio was 1.99) (Jung *et al.*, 2011). The results showed a higher number of SNPs achieved as compared with recent studies on blunt snout bream (Gao *et al.*, 2012). This could be explained that the Solexa/Illumina sequencing method used in this study, compared with 454 pyrosequencing, provides high-density markers, indicating a better approach which may be applied for large scale prediction of molecular markers. However, a validation for eliminating false positives and sequencing errors in this work needs to be performed on all potentially identified SNP markers. In facts, SNPs could be used as good markers in genomic mapping and finding candidate genes for quantitative trait loci selection (Liu and Cordes, 2004). Therefore, these putative SNPs identified in the present study would be priority candidates for marker development and be useful for further genetic and genomic studies such as genetic variation, population structure, conservation genetics, diversity analysis, and molecular assistant breeding on blunt snout bream.

The SNPs distribute in the immune-related unigenes obtained in this study provide basic information on the immunogenomics field in blunt snout bream. Interestingly, a higher number of SNPs distributed in genes involving in the immune responses and disease processing mechanisms (Table 1). Actually, a total of 111 (5.4%) unigenes including 600 SNPs were successfully annotated in the term 'immune system'. The results of number of SNPs identified from the blunt snout bream transcriptome profile in this work was lower than that from the *Mytilus chilensis* transcriptome, which had 20,306 SNPs in immune-related genes (Núñez-Acuña and Gallardo-Escárate, 2013). This difference may be a different applied method in selection of immune-related genes and in this study only genes associated with the term 'immune system' were investigated. In this study, the genes, A2M, TLR5 and C3 were found to be contained the highest number of SNPs. These genes were described playing important roles in the immune system of animals (Hayashi *et al.*, 2001; Magnadóttir *et al.*, 2005; Magnadottir, 2010; Zappia *et al.*, 2004). In relevance, a previous study reported that SNPs in A2M is related to the increasing risk for Alzheimer disease in human (Zappia *et al.*, 2004). This suggests that the SNPs identified distributing in the immune-related genes may play a crucial potential role involving in resistance or susceptibility to invasive pathogens of blunt snout bream. To our knowledge, this is the first comprehensive on the discovery of SNPs located in innate immune-related genes from the transcriptome profile of blunt snout bream. The SNPs found in this work would be basic information supporting to the immunogenomic researches and the developing molecular markers involving in resistance/susceptibility to invasive pathogens of blunt snout bream as well in future.

***Acknowledgements*:** The first author Ngoc Tuan Tran would like to thank the China Scholarship Council for providing

scholarship of doctoral program at Huazhong Agricultural University, Wuhan, Hubei Province, P.R. China. This work was financially supported by Modern Agriculture Industry Technology System Construction Projects of China titled as-Staple Freshwater Fishes Industry Technology System (No. CARS-46-05), National Natural Science Foundation of China (No. 31201988), Hubei Cooperation and Exchanges (No. 2013BHE006), Wuhan Engineering Research Center of Breeding for Wuchang Bream (No. 201302070510318), Excellent Youth Foundation of Hubei Scientific Committee (No. 2013CFA032) and Fundamental Research Funds for the Central Universities (No. 2013PY066 and 2014PY013).

REFERENCES

Beaumont, M.A. 2005. Adaptation and speciation: what can FST tell us? Trends Ecol. Evol. 20:435-440.

Collins, D.W. and T.H. Jukes. 1994. Rates of transition and transversion in coding sequences since the human-rodent divergence. Genomics 20:386-396.

de Boer, M.G.J., H. Jolink, C.J. Halkes, P.L. van der Heiden, D. Kremer, J.F. Falkenburg, E. van de Vosse and J.T. van Dissel. 2011. Influence of polymorphisms in innate immunity genes on susceptibility to invasive aspergillosis after stem cell transplantation. PLoS One 6:e18403.

Gao, Z., W. Luo, H. Liu, C. Zeng, X. Liu, S. Yi and W. Wang. 2012. Transcriptome analysis and SSR/SNP markers information of the blunt snout bream (*Megalobrama amblycephala*). PloS One 7:e42637.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644-652.

Hartel, C., D. Finas, P. Ahrens, E. Kattner, T. Schaible, D. Müller, H. Segerer, K. Albrecht, J. Möller and K. Diedrich. 2004. Polymorphisms of genes involved in innate immunity: association with preterm delivery. Mol. Hum. Reprod. 10:911-915.

Hayashi, F., K.D. Smith, A. Ozinsky, T.R. Hawn, C.Y. Eugene, D.R. Goodlett, J.K. Eng, S. Akira, D.M. Underhill and A. Aderem. 2001. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature 410:1099-1103.

He, C., L. Chen, M. Simmons, P. Li, S. Kim and Z. Liu. 2003. Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. Anim. Genet. 34:445-448.

He, L.-J., L.-K. Liao, J.-F. Yuan, H.-Y. Tang, Q. Wu and G.-W. Zhang. 2006. Pathological observation of bacterial septicemia in *Megalobrania amblycephala*. J. Southwest Agri. Univ. 3:035.

Helyar, S.J., J. Hemmer-Hansen, D. Bekkevold, M.I. Taylor, R. Ogden, M.T. Limborg, A. Cariani, G.E. Maes, E. Diopere and G.R. Carvalho. 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Mol. Ecol. Resour. 11:123-136.

Hinds, D.A., L.L. Stuve, G.B. Nilsen, E. Halperin, E. Eskin, D.G. Ballinger, K.A. Frazer and D.R. Cox. 2005. Whole-genome patterns of common DNA variation in three human populations. Science 307:1072-1079.

Jung, H., R.E. Lyons, H. Dinh, D.A. Hurwood, S. McWilliam and P.B. Mather. 2011. Transcriptomics of a giant freshwater prawn (*Macrobrachium rosenbergii*): de novo assembly, annotation and marker discovery. PloS One 6:e27938.

Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno and M. Hattori. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. 32:D277-D280.

Kongchum, P., Y. Palti, E.M. Hallerman, G. Hulata and L. David. 2010. SNP discovery and development of genetic markers for mapping innate immune response genes in common carp (*Cyprinus carpio*). Fish Shellfish Immunol. 29:356-361.

Kucuktas, H., S. Wang, P. Li, C. He, P. Xu, Z. Sha, H. Liu, Y. Jiang, P. Baoprasertkul and B. Somridhivej. 2009. Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. Genetics 181:1649-1660.

Li, S. and W. Cai. 2003. Genetic improvement of the herbivorous blunt snout bream (*Megalobrama amblycephala*). Naga 26:20-23.

Liu, Z. and J. Cordes. 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture 238:1-37.

Luikart, G., P.R. England, D. Tallmon, S. Jordan and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. 4:981-994.

Magnadóttir, B., S. Lange, S. Gudmundsdottir, J. Bøgwald and R. Dalmo. 2005. Ontogeny of humoral immune parameters in fish. Fish Shellfish Immunol. 19:429-439.

Magnadottir, B. 2010. Immunological control of fish diseases. Mar. Biotechnol. (NY) 12:361-379.

MAPRC (Ministry of Agriculture of the People's Republic of China). 2012. Chinese fisheries yearbook. Chinese Agricultural Press, Beijing, China.

Ming, J., J. Xie, P. Xu, X. Ge, W. Liu and J. Ye. 2012. Effects of emodin and vitamin C on growth performance, biochemical parameters and two HSP70s mRNA expression of Wuchang bream (*Megalobrama amblycephala* Yih) under high temperature stress. Fish Shellfish Immunol. 32:651-661.

Moriya, Y., M. Itoh, S. Okuda, A.C. Yoshizawa and M.

Kanehisa. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182-W185.

Morton, B.R., I.V. Bi, M.D. McMullen and B.S. Gaut. 2006. Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition. Genetics 172:569-577.

Núñez-Acuña, G. and C. Gallardo-Escárate. 2013. Identification of immune-related SNPs in the transcriptome of *Mytilus chilensis* through high-throughput sequencing. Fish Shellfish Immunol. 35:1899-1905.

Nielsen, M.E., L. Hoi, A. Schmidt, D. Qian, T. Shimada, J. Shen and J. Larsen. 2001. Is *Aeromonas hydrophila* the dominant motile *Aeromonas* species that causes disease outbreaks in aquaculture production in the Zhejiang Province of China? Dis. Aquat. Org. 46:23-29.

Ning, Z., A.J. Cox and J.C. Mullikin. 2001. SSAHA: a fast search method for large DNA databases. Genome Res. 11:1725-1729.

Picoult-Newberg, L., T.E. Ideker, M.G. Pohl, S.L. Taylor, M.A. Donaldson, D.A. Nickerson and M. Boyce-Jacino. 1999. Mining SNPs from EST databases. Genome Res. 9:167-174.

Salem, M., R.L. Vallejo, T.D. Leeds, Y. Palti, S. Liu, A. Sabbagh, C.E. Rexroad III and J. Yao. 2012. RNA-Seq identifies SNP markers for growth traits in rainbow trout. PLoS One 7:e36264.

SiFa, L. and C. WanQi. 2000. Two-way selective response of *Megalobrama amblycephala*. J. Fisheries of China 24:201-205.

Tran, N.T., Z.-X. Gao, J. Milton, L. Lin, Y. Zhou and W.-M. Wang. 2015a. Pathogenicity of *Aeromonas hydrophila* to blunt snout bream *Megalobrama amblycephala*. IJSRP 5:1-7.

Tran, N.T., Z.-X. Gao, H.-H. Zhao, S.-K. Yi, B.-X. Chen, Y.-H. Zhao, L. Lin, X.-Q. Liu and W.-M. Wang. 2015b. Transcriptome analysis and microsatellite discovery in the blunt snout bream (*Megalobrama amblycephala*) after challenge with *Aeromonas hydrophila*. Fish Shellfish Immunol. 45:72-82.

Tsao, W. 1960. A biological study of *Magalobrama amblycephala* and *M. terminalis* of Liang-tse Lake. Acta Hydrobiol. Sin. 1:57-78.

Vignal, A., D. Milan, M. SanCristobal and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34:275-306.

Wang, C., H. Wang, Y. Li and B. Liu. 2012a. Identification of a fructose-1, 6-bisphosphate aldolase gene and association of the single nucleotide polymorphisms with growth traits in the clam *Meretrix meretrix*. Mol. Biol. Rep. 39:5017-5024.

Wang, C., Y. You, H. Wang and B. Liu. 2012b. Genetic diversity of the sulfotransferase-like gene and one nonsynonymous SNP associated with growth traits of clam, *Meretrix meretrix*. Mol. Biol. Rep. 39:1323-1331.

Yu, Y., J. Wei, X. Zhang, J. Liu, C. Liu, F. Li and J. Xiang. 2014. SNP discovery in the transcriptome of white Pacific shrimp *Litopenaeus vannamei* by next generation sequencing. PloS One 9:e87218.

Zappia, M., I. Manna, P. Serra, R. Cittadella, V. Andreoli, A. La Russa, F. Annesi, P. Spadafora, N. Romeo and G. Nicoletti. 2004. Increased risk for Alzheimer disease with the interaction of MPO and A2M polymorphisms. Arch. Neurol. 61:341-344.

Zhou, Z., Z. Ren, H. Zeng and B. Yao. 2008. Apparent digestibility of various feedstuffs for bluntnose black bream *Megalobrama amblycephala* Yih. Aquaculture Nutrition 14:153-165.