Divergence and polymorphism analysis of IGF1Ra and IGF1Rb from orange-spotted grouper, *Epinephelus coioides* (Hamilton)

L. Guo, S. Yang, M.M. Li, Z.N. Meng and H.R. Lin

State Key Laboratory of Biocontrol, Institute of Aquatic Economic Animals and the Guangdong Province Key Laboratory for Aquatic Economic Animals, School of Life Sciences, Sun Yat-Sen University, Guangzhou, China

Corresponding author: Z.N. Meng
E-mail: mengzn@mail.sysu.edu.cn

ABSTRACT. Orange-spotted grouper (*Epinephelus coioides*) is one of the most important marine fish and has a high market value. The insulin-like growth factor type 1 receptor (IGF1R) is a component of the insulin-like growth factor signaling system, and demonstrates important roles during growth. Based on information from livestock, we used IGF1R as a candidate gene to survey single nucleotide polymorphisms. In the present study, the sequences of IGF1Ra and IGF1Rb from orange-spotted grouper were obtained from the genome sequences and their clustering in clades a and b, respectively, was confirmed by phylogenetic analysis. Fourteen critical amino acids underlying functional divergence were detected between the two clades, revealing the molecular basis of their functional differences. Nearly one-fourth (22 kbp) of the genomic sequence of IGF1Ra was sequenced in a mass cross population, and nucleotide diversity and linkage disequilibrium levels

were investigated. Nucleotide diversity was 0.00328 for π and 0.00344 for θ_w. The half decay of the squared allele-frequency correlation was 10,835 base pairs. Comparatively, the relatively high level of linkage and the significant deviation from neutrality-based codons in IGR1R showed that this gene was under selection. A site (KR269824.1:g. 63762C>T), located in the sixth intron, was significantly associated with eyeball diameter (P = 1.39 x 10^{-4}, Q-value: 2.33 x 10^{-2}), which accounted for 11.1% of phenotypic variance. These results highlight the important function of IGF1R in orange-spotted grouper and may be beneficial in the breeding of this species.

INTRODUCTION

Orange-spotted grouper (*Epinephelus coioides*), which belongs to the Serranidae family (Froese and Pauly, 2011), is treated as a model species and is used to determine the genetic and physiological features of grouper, including the phenomenon of sex reversal. Moreover, this species has been a popular stock in aquaculture, particularly in Southeast Asia, due to its robustness in overcrowded conditions, fast growth at high temperatures, relatively high price in the market, and the insufficient supply of wild captured fish (Pierre et al., 2008). Global production of this species, including aquaculture and capture production, has rapidly increased, reaching 9357 tons in 2014, with aquaculture production also growing fast in recent years (FAO, 2016). Although the reproduction of orange-spotted grouper has been fully controlled, the fry are collected directly or indirectly from their natural habitat, which increases the risk in fish culture due to the unstable germplasm (Pierre et al., 2008). With the rapid development of economic growth in Southeast Asia, extensive culture is not sufficient to meet the demand of the market. There is a demand for high-quality fry to improve the quality of broodstock. Marker-assisted breeding has become an effective method for selecting high-quality stock, and linkage disequilibrium mapping (LD-mapping), also known as association mapping or association analysis, has been used to dissect the relationship between genotype and phenotype (Thavamanikumar et al., 2011). LD-mapping utilizes historically accumulated crossover, resulting in high resolution, and can be applied to random populations and breeding populations, while family-based quantitative trait locus mapping is limited to family population. A unified mixed model has been introduced to account for multiple levels of relatedness, including kinship and patterns of population structure (Yu et al., 2006). This method has been widely used to determine the association between quantitative traits and genotypes of candidate genes (Thavamanikumar et al., 2011). In addition, we have used this method to detect the association between growth traits and genotypes in growth hormone-releasing hormone (GHRH), the GHRH receptor, and PACAP-related peptide/pituitary adenylate cyclase activating polypeptide type a (PRP-PACAPa), and type b (PRP-PACAPb) in this species (Guo et al., 2015a).

Due to its central importance in growth regulation, the insulin-like growth factor (IGF) signaling system plays an important role in determining growth patterns (Maures et al., 2002). The function of the IGF type 1 receptor (IGF1R) as an important member of this signaling system has been explored, particularly in humans. This gene is associated with

height in human (Fontenele et al., 2015). In mice, null mutants of this gene die invariably at birth due to respiratory failure and exhibit severe growth deficiency (Liu et al., 1993). Information on the important functions of this gene in growth and development is based on evidence from mammals, which suggests that this gene may be a target in agriculture. A large number of association analyses have been performed in livestock. This gene is associated with growth (Szewczuk et al., 2013) in cattle, reduced size (Hoopes et al., 2012) in dogs, body weight in Japanese quail (Moe et al., 2007), and early growth traits, carcass traits (Lei et al., 2008), and feed conversion ratio (Jin et al., 2014) in chicken. Accumulated evidence to date suggests that the major components of the IGF signaling system in teleost are structurally and functionally similar to those in mammals. Nevertheless, two subtypes of this gene, IGF1Ra and IGF1Rb, exist in teleost (Maures et al., 2002), which originated from gene duplication (Schlueter et al., 2006). Evidence from knockdown models in zebrafish showed that these two subtypes are required for proper embryonic growth, development, and survival (Schlueter et al., 2006), and diverge in their expression patterns (Maures et al., 2002; Li et al., 2014) and function, including early germline development (Schlueter et al., 2007) and the promotion of motor neuron innervation (Schlueter et al., 2006). Full-length cDNA sequences of both subtypes have also been cloned in Japanese flounder (*Paralichthys olivaceus*) (Nakao et al., 2002) and common carp (*Cyprinus carpio*). The full-length sequence of one of these subtypes has been cloned in orange-spotted grouper, although the clade has not been identified and the sequence has not been submitted to a gene bank (Kuang et al., 2005). This subtype has been proposed as subtype b following a primary search of the gene bank at the National Center for Biotechnology Information (NCBI). Unlike other genes in the IGF signaling system, this gene has not been targeted for association analyses in fish.

This study was performed to survey single nucleotide polymorphisms (SNPs) in the IGF1R gene and to investigate their association with growth traits. First, we obtained DNA and cDNA sequences of the IGF1Ra gene from the orange-spotted grouper genome sequences and the cDNA sequence of IGF1Rb from previous publication (Kuang et al., 2005). Second, the clades of these two IGF1R subtypes were confirmed and critical amino acid residues involved in their function were identified. Third, the IGF1Ra gene was selected for sequencing in a breeding population and nucleotide diversity, LD, and association analysis between genotypes and phenotypes in IGF1Ra were assessed. The estimated state of SNPs in the breeding population is expected to be meaningful for breeding in this species.

MATERIAL AND METHODS

Samples

The samples and phenotypic records studied have been described previously (Guo et al., 2015a). Briefly, the parents were captured in the South China Sea near Hainan Island (18°N, 109°E). First generation progenies were produced from a mass cross. Once the first generation progenies had matured, the second generation progenies were generated from a mass cross of 29 males and 12 females. The second generation fry, which numbered approximately 15,000, were fed according to the management practices of the fishery farm and cultured for 8 months in the same net cage in Haitang Bay of Sanya (Hainan Province, China). Additionally, 159 fry were randomly selected and 11 phenotypes were recorded, including body weight (BWT), maximum body depth, total length, standard length (SL), head length, body width, caudal

peduncle length, caudal peduncle depth, snout length, eyeball diameter (ED), and interorbital distance. The condition factor (K) was calculated using the following formula: $K = 100 \times BWT / SL^3$. Tissues were sectioned from the caudal fin of each individual and preserved in 95% ethanol at -20°C immediately thereafter. DNA was extracted and the quality was confirmed on 1% agarose gels and a UV spectrophotometer (Nanodrop2000/2000c, Thermo Scientific, Pittsburgh, PA, USA). All animal experiments were approved and conducted in accordance with the guidelines of the Animal Research and Ethics Committees of Sun Yat-Sen University.

Phylogenetic analysis of IGF1R

An IGF1R mRNA sequence in orange-spotted grouper was obtained from the previously published paper (Kuang et al., 2005), which was predicted to be subtype b. The genomic and mRNA sequences of presumed IGF1Ra (accession No.: KR269824.1) were obtained from the genome sequences of orange-spotted grouper (Zhang Y, unpublished data). To confirm the clade of the two genes, additional full length mRNA sequences were downloaded from the NCBI database, including IGF1Ra (zebrafish: AF400275.1, Japanese flounder: AB065098.1, and common carp: AY144591.1), IGF1Rb (zebrafish: AF400276, Japanese flounder: AB065099.1, and common carp: AY144592.1), IGF1R (human: X04434.1, mouse: AF056187.1, chicken: AJ223164.1, and Xenopus: AF055980.1), and insulin receptor (IR) (human: M10051.1), in which IR was treated as the outgroup. The translated protein sequences were aligned using MUSCLE and evolutionary trees were constructed using the MEGA6 software (Tamura et al., 2013). Neighbor-joining (NJ) and maximum likelihood (ML) methods were used with the default parameters.

To detect the divergent sites of the two IGF1R clades, the protein sequences of zebrafish IGF1Ra and IGF1Rb were used to search the database of non-redundant protein sequences at NCBI with an e-value set to 10^{-15}. In total, 250 of the closest related protein sequences of each search batch were downloaded. Sequences from teleost and those detailed above were selected for alignment, and a protein sequence from *Callorhinchus milii* was included as the outgroup (Table 1). NJ and ML trees were constructed as above. The DIVERGE 3.0 program (Gu and Vander Velden, 2002) was used to analyze functional divergence between the two clades of the IGF1R gene. DIVERGE estimates the ML of the theta (θ) Type-I and Type-II coefficients of functional divergence based on the occurrence of altered selective constraints and radical shifts of physicochemical properties, respectively. The θ value indicates the extent of functional divergence, ranging from 0 (no functional divergence) to 1 (maximum divergence). θ (Type-I) was calculated using the model-free method and θ (Type-II) was calculated using the simplified ML method. Statistical testing to determine whether θ was significantly larger than 0, was based on the estimate (θ) and its standard error (SE), in which the Z-score (θ/SE) follows a normal distribution. Amino acid positions with posterior probabilities (Q) greater than 0.8 were selected as candidates. The candidates may have functional roles in the selective divergence of subclasses and may represent the residues responsible for subclass-specific activity.

Target sequencing and SNP determination

Four pairs of primers (Table 2) were previously designed (Guo et al., 2015b). DNA amplification, sequencing and SNP calling were performed according to previous methods (Guo et al., 2015a).

Table 1. List of investigated insulin like growth factor 1 receptor (IGF1R) sequences.

Species	Subtype	GenBank and RefSeq
Fundulus heteroclitus	a (isoform X2)	gi|831553915|ref|XP_012729198.1
	b	gi|831539861|ref|XP_012724167.1
Poecilia formosa	a	gi|617378672|ref|XP_007545795.1
	b	gi|617453066|ref|XP_007569286.1
Austrofundulus limnaeus	a	gi|928021718|ref|XP_013858973.1
	b	gi|928065420|ref|XP_013882332.1
Stegastes partitus	a	gi|657554530|ref|XP_008281734.1
	b	gi|657557982|ref|XP_008283098.1
Paralichthys olivaceus	a	gi|18150108|dbj|BAB83669.1
	b	gi|18150110|dbj|BAB83670.1
Takifugu rubripes	a	gi|768925410|ref|XP_003967106.2
	b	gi|410912006|ref|XP_003969481.1
Haplochromis burtoni	a	gi|554845552|ref|XP_005933117.1
	b	gi|554804284|ref|XP_005913097.1
Pundamilia nyererei	a	gi|548355601|ref|XP_005729969.1
	b	gi|548350634|ref|XP_005727628.1
Maylandia zebra	a	gi|498994601|ref|XP_004553288.1
	b	gi|498930890|ref|XP_004539804.1
Clupea harengus	a	gi|831310809|ref|XP_012687827.1
	b	gi|831287087|ref|XP_012674886.1
Danio rerio	a	gi|23308649|ref|NP_694500.1
	b	gi|23308651|ref|NP_694501.1
Cyprinus carpio	a	gi|33355376|gb|AAN52151.1
	b	gi|33355378|gb|AAN52152.1
Oryzias latipes	a (isoform X2)	gi|432861730|ref|XP_004069709.1
	a (isoform X3)	gi|765122778|ref|XP_011474451.1
	a (isoform X1)	gi|765122775|ref|XP_011474450.1
	b	gi|765116465|ref|XP_011492659.1
Oreochromis niloticus	a	gi|542190399|ref|XP_003440646.2
	b (isoform X2)	gi|908429027|ref|XP_013132306.1
	b (isoform X1)	gi|542213912|ref|XP_005447746.1
Poecilia_reticulate	a	gi|658855928|ref|XP_008410609.1
Xiphophorus maculatus	a	gi|941787339|ref|XP_014325343.1
Paralichthys adspersus	a	gi|215434963|gb|ACJ66864.1
Sparus aurata	a	gi|747174769|gb|AJE25612.1
Esox lucius	a	gi|742192444|ref|XP_010891000.1
Nothobranchius furzeri	a	gi|319769163|gb|ADV72543.1
Salmo salar	a	gi|929256466|ref|XP_014004580.1
	a (isoform X2)	gi|929123899|ref|XP_013979595.1
	a (isoform X1)	gi|929123897|ref|XP_013979594.1
Poecilia reticulata	b	gi|658842658|ref|XP_008403720.1
Larimichthys crocea	b (1)	gi|808882895|gb|KKF30174.1
Larimichthys crocea	b (2)	gi|734608293|ref|XP_010731390.1
Scophthalmus maximus	b	gi|3219622|emb|CAA12278.1
Cynoglossus semilaevis	b	gi|657747460|ref|XP_008309033.1
Xenopus laevis		gi|147904186|ref|NP_001081734.1
Gallus		gi|45384296|ref|NP_990363.1
Mus musculus		gi|112983656|ref|NP_034643.2
Homo sapiens		gi|4557665|ref|NP_000866.1
Callorhinchus milii		gi|632938828|ref|XP_007906561.1
IR Homo sapiens		gi|307070|gb|AAA59174.1

High-fidelity DNA polymerase, PrimeSTAR GXL DNA Polymerase (TaKaRa, Dalian, China), was used for amplifications. The amplicons were sequenced on an ion torrent personal genomic machine using chip 318. High-quality SNPs were identified using the following criteria: 1) base quality levels, map quality levels, and SNP quality levels should be greater than or equal to 20 (Phred scale); 2) SNP coverage level should be greater than or equal to 10-fold; and 3) a mutation with a ratio of variant coverage (variant/total) from 10 to 90% should be defined as a heterozygous variant. Physical coverage was defined as the percentage of sites with a mean coverage depth above 10.

Table 2. Primers specific for the IGF1-Ra gene (four fragments).

Description	Forward 5' to 3'	Reverse 5' to 3'	Length (bp)
IGF1Ra.H1	CCCTGTGAATCTGGACGTG	TTCCCCAAAACCCCAAACA	5231
IGF1Ra.H2	AGTATGCCCACAACACTGCAAG	CGACTCCTTGTAGTAGACTAT	5313
IGF1Ra.H3	CACAGAGTTTGACGGGCAGGAT	CACGCACAACATTACAACAAGCATT	6176
IGF1Ra.H4	TGTATGTCCCAGATGAGTGG	CCAAGAGTCCTCCTTCCAT	5581

Association analysis between genotypes and phenotypes and estimates of nucleotide diversity

Levels of LD between each pair of SNPs with a minor allele frequency (MAF) >10% were estimated as squared allele-frequency correlation (r^2) using the Haploview 4.2 program (Barrett et al., 2005). The LD decay with distance was evaluated using a mutation-recombination drift model (Sved, 1971). Association analysis was performed using a mix linear model, which was made available through the TASSEL software program version 5.1.0 (Bradbury et al., 2007), in which kinship and population structure were considered to improve the true positive rate. SNPs with a MAF of >5% and a missing rate of <20% were selected, and the pairwise kinship coefficients and population structure matrix were previously estimated (Guo et al., 2015a). Multiple comparisons were adjusted using the false discovery ratio (Q) (Storey and Tibshirani, 2003). The variation explained by each SNP (R^2) was estimated as: $R^2 = SS_t / SS_T \times 100\%$, where SS_t is the variance explained by the marker and SS_T is the total variance. The gene action model was qualified using the ratio of dominance (d) to additive (a) effects (Tian et al., 2014).

A codon-based Z-test of the selection of all sequences was performed using the MEGA6 software (Tamura et al., 2013). The test statistic of the difference in the number of synonymous and nonsynonymous substitutions per site was calculated using the Nei-Gojobori method (Nei and Gojobori, 1986) and the variance of the difference was computed using the bootstrap method (1000 replicates). Haplotypes and missing genotypes were inferred using the PHASE 2.1.1 software (Stephens et al., 2001). Varying rates of recombination were applied, and burn-in and iteration quantities were both set to 100. Nucleotide diversity levels were estimated as p (Nei, 1987) and q_w (Watterson, 1975), and allele frequency spectra was estimated as Tajima's D using the DnaSP software version 5.10 (Librado and Rozas, 2009).

RESULTS

The genomic sequence of the putative IGF1Ra was obtained from the orange-spotted grouper genome sequence (Zhang Y, unpublished data) and the gene was found to contain 21 exons (Figure 1). Phylogenetic analysis supported the two clades of IGF1R in teleost with 100% of bootstrap replicates using both the NJ and ML methods. The IGF1R gene obtained

from the orange-spotted grouper genome sequence was clustered in subtype a, and the previously reported IGF1R gene was clustered in subtype b, which confirmed the clade of the two IGF1R in orange-spotted grouper (Figure 2). Previous physiological experiments in zebrafish have demonstrated that the two subtypes have different roles during development (Schlueter et al., 2006). We collected 47 IGF1R protein sequences from teleost in order to determine the molecular basis of their functional divergence, which included almost all of the sequences from teleost. Sequences from teleost were grouped into two clades with 100% of the bootstrap replicates in both methods (NJ and ML), and all of these gene subtypes were clustered into the corresponding clade independently of the method used (NJ and ML). Figure 3 shows the cluster using the NJ method. The functional divergence between the two subtypes was estimated using the DIVERGE 3.0 program (Gu and Vander Velden, 2002) (Table 3). Type I functional divergence showed $\theta = 0.294 \pm 0.031$, which was significantly greater than 0 ($P < 0.00001$). Thus, the functional divergence at the primary structure level of protein between the two clades was significant, and 14 residues had a stringent threshold of posterior probability higher than 0.80. Type II functional divergence showed $\theta = 0.0023 \pm 0.036$, which was not significant according to a statistical test.
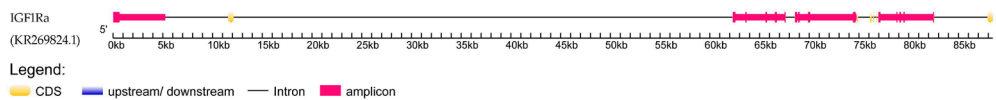

Figure 1. Illustration of the insulin-like growth factor a (IGF1Ra) gene and the position of the four amplicons. The IGF1Ra gene contains 21 exons and 87.9 kbps from the start codon to the stop codon. This depiction shows the structure of the entire gene, including 551 bp upstream of the gene. The amplicons covered the most exons, except for the 2nd, 12th, 13th, 14th, 15th, and 21st exons. The four amplicons (IGR1R.H1, IGR1R.H2, IGR1R.H3, and IGR1R.H4), excluding the length of the primers contained 5231, 5313, 6176, and 5581 base pairs, respectively, in which the first amplicon (H1) contained 551 bp upstream of the start codon (pink).
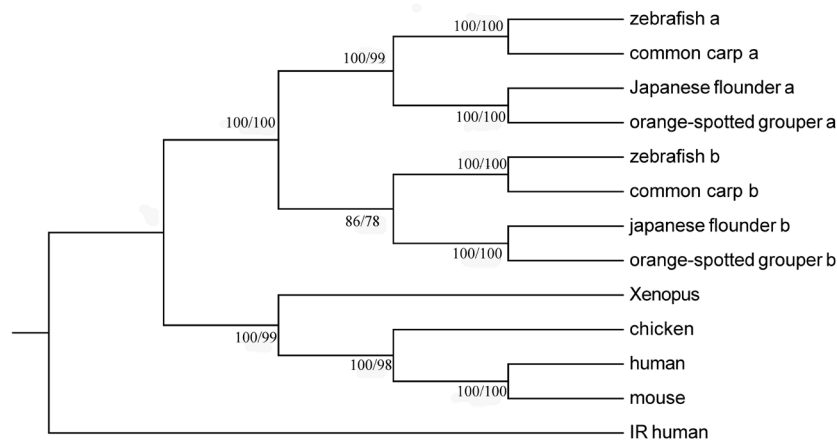

Figure 2. Phylogenetic relationships of the IGF1R gene inferred using the neighbor-joining (NJ) and maximum likelihood (ML) methods. Numbers at the nodes indicate the bootstrap values (%), in which the numbers present before the slash indicate the values (%) determined using the NJ method and those after the slash indicate the values using the ML method. The two IGF1R genes from orange-spotted grouper were clustered with subtype a and b, respectively, which confirmed that the corresponding genes were IGF1Ra and IGF1Rb.
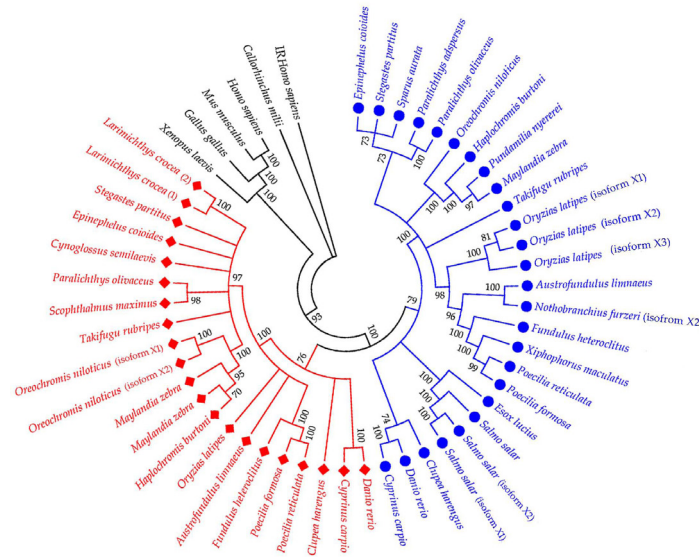
Figure 3. Phylogenetic relationships of IGF1R inferred using the NJ method. Numbers at the nodes indicate the bootstrap value (%). The blue branch indicates the gene clade IGF1Ra and the red branch indicates the gene clade IGF1Rb.

Table 3. Analysis of functional divergence between IGF1Ra and IGF1Rb.

Functional divergence	Coefficient θ ± SE	Critical amino acid sites
Type I	0.2944 ± 0.031 (P < 0.00001)	155, 174, 187*, 223, 238, 287, 290*, 320, 331, 339*, 344, 357, 366, 370, 412*, 516, 565, 590, 685, 720, 765*, 767*, 828, 845, 903*, 917, 918*, 1089, 1143, 1287*, 1288, 1290*, 1292, 1298*, 1299*, 1300*, 1319, 1320, 1321, 1335*, 1337
Type II	0.0023 ± 0.036 (P = 0.47)	

P, posterior probability values; SE, standard error. Critical amino acid sites related to functional divergence with posterior possibility (Q) > 0.50 (*Q > 0.80) are listed. Numbering refers to the human IGF1R protein sequence (GenBank and RefSeq: gi|4557665|ref|NP_000866.1).

To investigate the nucleotide diversity, the level of LD, and association between phenotypes and genotypes, we sequenced four fragments (IGF1Ra.H1, IGF1Ra.H2, IGF1Ra. H3, and IGF1Ra.H4) within the IGF1Ra gene in the breeding population. The positions of the four amplicons are shown in Figure 1. The lengths of the four amplicons were 5231, 5313, 6176, and 5581 bp, respectively, excluding the length of the primers, which covered approximately 25% of the gene. Amplicon sequencing obtained 125 ± 59, 124 ± 55, 120 ± 57, and 122 ± 64 (mean ± SD)-fold mean coverage depth for IGF1Ra.H1, IGF1Ra.H2, IGF1Ra.H3, and IGF1Ra.H4, respectively (Figure 4a). Furthermore, 95, 96, 98, and 93% of individuals obtained at least 80% physical coverage for each amplicon, respectively (Figure 4b). In addition, 485 high-quality SNPs were detected, including three nonsynonymous mutations. The SNP frequency was 21.8 per kilo base pair. Moreover, 164 SNPs exhibited MAF >5%, and 131 SNPs exhibited MAF >10%, which was used to assess the association analysis and level of LD, respectively.
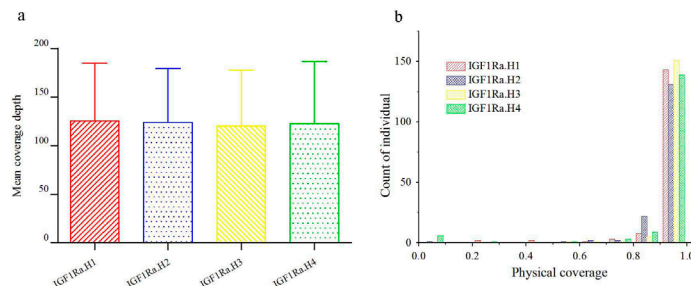
Figure 4. Statistics of the mean coverage depths and physical coverage of the four amplicons. **a.** Mean coverage depths for each individual and amplicon were calculated and are reported as the mean ± SD. The mean coverage depth was 125-fold for all amplicons. **b.** Physical coverage for each individual. The amplicons were calculated and nearly all of the fragments achieved 80% physical coverage.

Nucleotide diversity was evaluated, and the levels of the covered regions of the gene were 0.00328 for π and 0.00344 for θ_w (Table 4). However, the levels of synonymous sites were 54 and 24 times those of non-synonymous sites for π and θ_w, respectively. Although Tajima's D values were insignificantly negative, the codon-based test of selection was significant (P = 0.043, statistic: -2.046). The average value of r^2 was 0.45, and the half decay of r^2 was 10,835 bp. A site (KR269824.1:g. 63762C>T), located in the sixth intron, was significantly associated with ED (P = 1.39 x 10^{-4}, Q-value: 2.33 x 10^{-2}), which accounted for 11.1% of the phenotypic variance. The MAF of this site was 31.8%. This marker-trait association was consistent with a partial to complete dominant effect.

Table 4. Nucleotide polymorphisms in the IGF1-Ra gene (KR269824.1).

Description	No. of sites	No. of polymorphic sites	Nucleotide diversity		Tajima's D[c]
			π	θ_w	
Synonymous	581.54	22	0.00488	0.00597	-0.47393
Silent[a]	20,377.54	483	0.00358	0.00374	-0.13069
Non-synonymous	1941.46	3	0.00009	0.00024	-0.92850
Total gene[b]	22319	486	0.00328	0.00344	-0.00344

[a]Silent substitutions were considered to be those in the coding and noncoding regions. [b]Total number of sites of the fragment was assessed. [c]Quality of Tajima's D for all types was not significant (P > 0.10).

DISCUSSION

In this study, the genomic and mRNA sequences of the IGF1Ra gene were obtained, and the clades of the two cDNA sequences in orange-spotted grouper were confirmed. Based on knowledge of their functional divergence from physiological experiments, we collected almost all of the protein sequences from teleost. Fourteen critical amino acids were detected based on the different evolutionary rates, which may explain the physical basis of the functional divergence at the primary structure level of the protein. Due to the important functions of the IGF signaling system, IGF1Ra was selected as a target gene, and nucleotide diversity was investigated in the breeding population. The level of LD within the gene was estimated and found to be relatively low. The associated site (KR269824.1:g. 63762C>T) with ED was detected. Because little is known about the genetics of this species, evaluation of the nucleotide diversity and level of LD may aid in breeding practices.

The level of nucleotide diversity in the gene regions was similar to that in GHRH, GHRH receptor, PRP-PACAPa, and PRP-PACAPb, which were analyzed based on the same population (Guo et al., 2015a). Furthermore, a region that included the IGF1Ra gene was inferred to be under selection. First, the large difference in nucleotide diversities between non-synonymous and synonymous sites showed that a mutation in the non-synonymous sites was swept. Second, a codon-based neutrality test indicated that variation in IGF1Ra significantly deviated from neutrality and was under purifying selection. Third, the half decay of r^2 was much larger than that in the GHRH, PRP-PACPAa, PRP-PACAPb, and GHRHR regions, which was limited within 1 kbp (Guo et al., 2015a). In practice, the culture of this species is difficult due to the occurrence of serious diseases (Nagasawa and Cruz-Lacierda, 2004). Usually, diseases can result in the death of fish batches, and until recently, the scale of the culture was small. Thus, these factors may enlarge the effect of drift and the strength of selection. The broodstock was captured from the wild ocean, and the wild population is overfished (IUCN, 2012). These two factors may contribute to selection. However, confirmation of selection and driving factors require further evidence.

A significant association between the SNP KR269824.1:g. 63762C>T and ED was detected. However, this evidence was not strong. According to Yan et al. (2011), the statistical power was only 26% with a false positive rate of 1.39×10^{-4}. Thus, this association was not sufficiently convincing within the background of the scale of the entire genome. Although no significant association was detected between harvest traits and genotypes, the identification of effective markers for detection should be encouraged. Orange-spotted grouper is an important aquaculture species, particularly in Southeast Asia. With economic development, the culture style is expected to be sustainable and environmentally friendly. Survey and utilization of the germplasm resource of this species is at the preliminary stage. Although this species is widely distributed in the open sea, resource investigations are limited within small regions, mainly in Southeast Asia (Antoro et al., 2006; Wang et al., 2009; Meng et al., 2012). In recent years, next generation sequencing (NGS) has revolutionized nucleotide sequencing (Helyar et al., 2011). The construction of genetic linkage maps for this species and association analyses between genes (GHRH, PRP-PACAPa, PRP-PACAPb, and GHRHR) and growth traits (Guo et al., 2015a) have benefited from the improvement of NGS. With the aid of NGS, whole genomic association has been successfully applied to search for powerful markers (Campbell et al., 2014). With this expectation, the next phase of investigation on orange-spotted grouper will be based on the whole genome considering large scale population resources.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Antoro S, Na-Nakorn U and Koedprang W (2006). Study of genetic diversity of orange-spotted grouper, *Epinephelus coioides*, from Thailand and Indonesia using microsatellite markers. *Mar. Biotechnol. (NY)* 8: 17-26.

Barrett JC, Fry B, Maller J and Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263-265.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, et al. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635.

Campbell NR, LaPatra SE, Overturf K, Towner R, et al. (2014). Association mapping of disease resistance traits in rainbow trout using restriction site associated DNA sequencing. *G3 (Bethesda)* 4: 2473-2481.

FAO (2016). Global Production Statistics 1950-2014. Food and Agriculture Organization. Available at http://www.fao.org/figis/servlet/SQServlet?file=/work/FIGIS/prod/webapps/figis/temp/hqp_3045327508054819982.xml&outtype=html. Accessed July 17, 2016.

Froese R and Pauly D (2011). FishBase: www.fishbase.org, version. Accessed on 10 April 2013.

Fontenele EG, Moraes ME, d'Alva CB, Pinheiro DP, et al. (2015). Association Study of GWAS-Derived Loci with Height in Brazilian Children: Importance of MAP3K3, MMP24 and IGF1R Polymorphisms for Height Variation. *Horm. Res. Paediatr.* 84: 248-253.

Gu X and Vander Velden K (2002). DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics* 18: 500-501.

Guo L, Xia J, Yang S, Li M, et al. (2015a). GHRH, PRP-PACAP and GHRHR Target Sequencing via an Ion Torrent Personal Genome Machine Reveals an Association with Growth in Orange-Spotted Grouper (*Epinephelus coioides*). *Int. J. Mol. Sci.* 16: 26137-26150.

Guo L, Xia J, Yang S, Li M, et al. (2015b). Rapid detection of SNPs in candidate genes regulating the growth of orange-spotted grouper, Epinephelus coioides (Hamilton, 1822), using semiconductor sequencing. *Aquac. Res., online* (DOI: 10.1111/are.12897).

Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, et al. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol. Ecol. Resour.* 11 (Suppl 1): 123-136.

Hoopes BC, Rimbault M, Liebers D, Ostrander EA, et al. (2012). The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. *Mamm. Genome* 23: 780-790.

IUCN (2012). IUCN Red List of Threatened Species. Version 2012.2. www.iucnredlist.org. Accessed on 10 April 2013.

Jin S, Chen S, Li H, Lu Y, et al. (2014). Associations of polymorphisms in GHRL, GHSR, and IGF1R genes with feed efficiency in chickens. *Mol. Biol. Rep.* 41: 3973-3979.

Kuang YM, Li WS and Lin HR (2005). Molecular cloning and mRNA profile of insulin-like growth factor type 1 receptor in orange-spotted grouper, *Epinephelus coioides. Acta Biochim. Biophys. Sin. (Shanghai)* 37: 327-334.

Lei M, Peng X, Zhou M, Luo C, et al. (2008). Polymorphisms of the IGF1R gene and their genetic effects on chicken early growth and carcass traits. *BMC Genet.* 9: 70.

Li J, Wu P, Liu Y, Wang D, et al. (2014). Temporal and spatial expression of the four Igf ligands and two Igf type 1 receptors in zebrafish during early embryonic development. *Gene Expr. Patterns* 15: 104-111.

Librado P and Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

Liu JP, Baker J, Perkins AS, Robertson EJ, et al. (1993). Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r). *Cell* 75: 59-72.

Maures T, Chan SJ, Xu B, Sun H, et al. (2002). Structural, biochemical, and expression analysis of two distinct insulin-like growth factor I receptors and their ligands in zebrafish. *Endocrinology* 143: 1858-1871.

Meng ZN, Yang S, Fan B, Wang L, et al. (2012). Genetic variation and balancing selection at MHC class II exon 2 in cultured stocks and wild populations of orange-spotted grouper (*Epinephelus coioides*). *Genet. Mol. Res.* 11: 3869-3881.

Moe HH, Shimogiri T, Kamihiraguma W, Isobe H, et al. (2007). Analysis of polymorphisms in the insulin-like growth factor 1 receptor (IGF1R) gene from Japanese quail selected for body weight. *Anim. Genet.* 38: 659-661.

Nagasawa K and Cruz-Lacierda ER (2004). Diseases of cultured groupers. Southeast Asian Fisheries Development Center, Aquaculture Department Iloilo, Philippines.

Nakao N, Tanaka M, Higashimoto Y and Nakashima K (2002). Molecular cloning, identification and characterization of four distinct receptor subtypes for insulin and IGF-I in Japanese flounder, *Paralichthys olivaceus. J. Endocrinol.* 173: 365-375.

Nei M (1987). *Molecular evolutionary genetics*. Columbia University Press: New York, NY, USA

Nei M and Gojobori T (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3: 418-426.

Pierre S, Gaillard S, Prevot-D'Alvise N, Aubert J, et al. (2008). Grouper aquaculture: Asian success and Mediterranean trials. *Aquat. Conserv.* 18: 297-308.

Schlueter PJ, Royer T, Farah MH, Laser B, et al. (2006). Gene duplication and functional divergence of the zebrafish insulin-like growth factor 1 receptors. *FASEB J.* 20: 1230-1232.

Schlueter PJ, Sang X, Duan C and Wood AW (2007). Insulin-like growth factor receptor 1b is required for zebrafish primordial germ cell migration and survival. *Dev. Biol.* 305: 377-387.

Stephens M, Smith NJ and Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68: 978-989.

Storey JD and Tibshirani R (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 100: 9440-9445.

Sved JA (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Popul. Biol.* 2: 125-141.

Szewczuk M, Zych S, Wójcik J and Czerniawska-Piątkowska E (2013). Association of two SNPs in the coding region of the insulin-like growth factor 1 receptor (IGF1R) gene with growth-related traits in Angus cattle. *J. Appl. Genet.* 54: 305-308.

Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729.

Thavamanikumar S, Tibbits J, McManus L, Ades P, et al. (2011). Candidate gene-based association mapping of growth and wood quality traits in *Eucalyptus globulus* Labill. *BMC Proc.* 5: O15.

Tian J, Chang M, Du Q, Xu B, et al. (2014). Single-nucleotide polymorphisms in PtoCesA7 and their association with growth and wood properties in *Populus tomentosa. Mol. Genet. Genomics* 289: 439-455.

Wang J, Guo F, Ding S and Wang J (2009). Genetic diversity of different geographical stocks of *Epinephelus coioides* by microsatellite DNA. *Mark. Sci.* 11: 60-64.

Watterson GA (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7: 256-276.

Yan J, Warburton M and Crouch J (2011). Association Mapping for Enhancing Maize (*Zea mays* L.) Genetic Improvement. *Crop Sci.* 51: 433-449.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38: 203-208.