

缢蛏 *IGFBP* 基因结构及生长性状相关 SNP 筛选

谢淑媚, 牛东红*, 阮海灯, 王泽, 王飞, 陈世娥, 李家乐

(上海海洋大学农业部淡水水产种质资源重点实验室, 上海 201306)

摘要: 类胰岛素生长因子结合蛋白(IGFBP)是 IGF 系统的一部分,主要参与 IGF 的运输、定位和生物活性调节。本研究采用 RACE 技术和长 PCR 技术,克隆了缢蛏 *IGFBP* 基因的 cDNA 和 DNA 全长序列,应用荧光定量 PCR 技术分析了缢蛏不同发育时期和不同组织中 *IGFBP* mRNA 的表达特征,并进一步筛选了 *IGFBP* 基因与生长性状相关的 SNP 位点。序列分析表明,缢蛏 *IGFBP* cDNA 序列全长 631 bp,包括 5'端非编码区 60 bp,3'端非编码区 136 bp 和开放阅读框 435 bp,编码 144 个氨基酸。该基因含有保守的 IGFBP-N 端,包含 12 个半胱氨酸残基,其中 1~18 个氨基酸为信号肽,属于分泌型蛋白。*IGFBP* DNA 全长 3 122 bp,其中包含 1 个内含子(2 687 bp)和 2 个外显子(200 和 235 bp)。荧光定量 PCR 结果显示,*IGFBP* mRNA 在消化腺组织中表达量最高;在缢蛏的稚贝期,*IGFBP* mRNA 呈现高表达,而在其他发育时期表达量低。在 *IGFBP* 基因中筛选到 4 个 SNP 位点,其中 1 个 SNP 位点与缢蛏的壳长和体重呈显著相关。

关键词: 缢蛏;类胰岛素生长因子结合蛋白(IGFBP);基因表达;单核苷酸多态性(SNP);关联分析

中图分类号: Q 785; S 966.1

文献标志码: A

缢蛏(*Sinonovacula constricta*)俗称蛏子,属于双壳纲(Bivalvia)、帘蛤目(Veneroidea)、竹蛏科(Solenidae)、缢蛏属(*Sinonovacula*),广泛分布于中国、日本和朝鲜等国的沿海地区,埋栖生活,生长快,味道鲜美,营养丰富,具有较高的经济价值,是中国四大海产贝类之一^[1]。近年来,随着缢蛏的人工繁育技术日趋成熟,养殖产业正在逐步形成,对生长快,养殖周期短的缢蛏品种的需求度越来越高,因此,研究与缢蛏生长发育机制相关的基因,认识其生长调控的分子机制,对缢蛏的良种选育、提高其养殖性能和推动养殖业的发展具有重要意义。

类胰岛素生长因子系统(insulin-like growth factor system,IGFs)主要由 IGF 配体(IGF-I,IGF-II)、IGF 受体(IGF receptor,IGFR-I 和 IGFR-II)及 IGF 结合蛋白(insulin-like growth factor binding protein,IGFBP)3 部分组成,该系统主要介导垂体

生长激素(growth factor,GH)的促生长作用,与脊椎动物的生长发育密切相关^[2]。在哺乳动物中有 6 个 IGFs,它们在 IGFs 中参与 IGF 的运输和生物活性调节^[3-5],对 IGF 配体功能的发挥具有重要作用。

IGFs 的作用机制在脊椎动物中已有大量研究,但在贝类中的研究极少。研究利用人类重组类胰岛素生长因子 1(human recombinant insulin-like growth factor 1,hrIGF1)注射马氏珠母贝(*Pinctada fucata*),结果表明,实验组类胰岛素多肽受体(insulin-related peptide receptor,IRR)的表达量显著提高,壳基质蛋白基因的表达量显著改变,推测贝类中存在某种激素调节通路,能调控贝类的生长发育^[6]。研究发现,马氏珠母贝中存在的胰岛素相关多肽(molluscan insulin-related peptides,MIPs)与其生长、增殖和代谢有关^[7]。

收稿日期:2014-12-24 修回日期:2015-03-13

资助项目:国家“八六三”高技术研究发展计划(2012AA10A400);国家自然科学基金(31472278);上海知识服务平台(ZF1206)

通信作者:牛东红,E-mail:dhniu@shou.edu.cn

此外,研究表明软体动物中的胰岛素相关多肽主要通过调控蛋白质和糖类的代谢,进而调节软体动物内脏和壳的生长发育^[8]。近年,在杂色鲍(*Haliotis diversicolor*)中,进行了IGFBP相关蛋白IGFBP7的研究,结果表明,IGFBP7对杂色鲍的幼虫附着变态及成体免疫过程都发挥着重要的作用^[9]。在虾夷扇贝(*Patinopecten yessoensis*)中克隆获得了IGFBP基因的全长序列,研究表明其在虾夷扇贝发育的各时期和在成体的各组织中均广泛表达,且基因中存在1个与生长性状相关的SNP^[10]。越来越多的研究结果表明,IGFs与贝类的生长发育相关,加强对贝类IGFs的研究,有助于为贝类的生长发育以及变态的分子调节机制的研究提供基础支持。

缢蛭在我国分布广泛,一年即可性成熟,为研究贝类的生长机制提供了良好的实验条件。目前,IGFs在缢蛭中还未见报道,本研究采用RACE和qRT-PCR等技术探索了缢蛭IGFBP基因的结构特征及各发育时期和成体各组织的相对表达特征,筛选了与生长相关的SNP位点,揭示该基因与缢蛭生长的相关性。研究结果为缢蛭IGFBP基因功能的研究及其生长发育的分子调控机制的研究提供理论基础,丰富了双壳贝类IGFBP基因信息,为缢蛭良种选育过程中候选基因的进一步筛选提供参考。

1 材料与方 法

1.1 实验材料

实验所用缢蛭均在浙江省宁海市长街镇岳井洋缢蛭良种繁殖场采集。随机选取同一批繁殖的240个1龄缢蛭进行编号,测量其壳长(53.02 ± 2.27) mm、壳宽(12.41 ± 0.91) mm、壳高(17.55 ± 0.91) mm和体质量(8.09 ± 1.23) g,取外套膜置于无水乙醇中,常温保存,用于筛选生长相关SNP位点。挑选6个健康的1龄缢蛭,分别取外套膜、水管、斧足、鳃、消化腺和性腺组织,放在装有RNastore的1.5 mL RNase-free离心管中,4℃保存。缢蛭早期发育阶段的样本分装在装有RNastore的1.5 mL RNase-free离心管中,4℃保存。低温运回实验室后,抽除RNastore,液氮预冷,保存于-80℃备用。

1.2 总RNA和基因组DNA的提取

参照RNAPrep pure动物组织总RNA提取试

剂盒(Tiagen)说明书提取总RNA,1%琼脂糖凝胶电泳检测RNA的完整性,用NanoDrop 2000C检测RNA的浓度和OD₂₆₀/OD₂₈₀的比值,保存于-80℃备用;利用海洋动物组织基因组DNA提取试剂盒(Tiagen)提取DNA,1%琼脂糖凝胶电泳检测DNA的完整性,用NanoDrop 2000C测定DNA的纯度和浓度,保存于-20℃备用。

1.3 IGFBP基因cDNA模板的合成

参照BD SMARTTM RACE cDNA Amplification Kit User Manual(Clontech)试剂盒操作要求合成5'-RACE-Ready cDNA,根据3'-Full RACE Core Set with PrimeScriptTM RTase(TaKaRa)试剂盒使用说明合成3'-RACE-Ready cDNA。用PrimeScriptTM RT Master Mix(Perfect Real Time)进行反转录,所得的cDNA作为荧光定量PCR的模板,-20℃保存备用。

1.4 IGFBP cDNA片段的获得

在缢蛭消化腺cDNA文库中检索到IGFBP片段的序列^[11],根据该序列设计上、下游引物并扩增,引物信息见表1,扩增体系为2×Taq PCR Master Mix 10 μL;F1和R1各0.6 μL(10 μmol/L);cDNA模板2 μL和RNase-free water 6.8 μL。反应条件为94℃ 2 min,94℃ 30 s,58℃ 30 s,72℃ 30 s,35个循环,72℃ 10 min,4℃保存。扩增产物经1%琼脂糖凝胶电泳验证后割胶回收(Axygen),并按说明书进行纯化,将回收产物连接至PGEM-T easy载体(Promega),16℃连接16 h,热激转化DH5α大肠杆菌(Tiagen),挑选6个阳性克隆送交上海生工生物工程有限公司测序。

1.5 cDNA全长及基因组序列的扩增

根据获得的cDNA中间片段,分别设计用于3'和5'端扩增的引物P1和P2。以稀释10倍后的cDNA为模板,用正向引物P2和Outer Primer进行3'-RACE扩增,体系为cDNA 2 μL,1×cDNA dilution buffer II 3 μL,P2和Outer Primer各1 μL,10×LA PCR buffer II(Mg²⁺ Free) 2 μL,MgCl₂(25 mmol/L) 1.5 μL,LA Taq(5 U/μL) 0.175 μL,RNase-free water 14.375 μL;条件为94℃ 2 min,94℃ 30 s,60℃ 30 s,72℃ 1 min,35个循环,72℃ 10 min。利用反向引物P1和UPM进行5'-RACE扩增,体系包括:cDNA 1.25 μL,10×BD Advantage 2 PCR buffer 2.5 μL,dNTP Mix(10 mmol/L) 0.5 μL,50×BD

Advantage 2 polymerase mix 0.5 μL , UPM (10 \times) 引物 2.5 μL , P1 引物 0.5 μL , RNase-free water 17.25 μL ; 条件为 94 $^{\circ}\text{C}$ 2 min, 94 $^{\circ}\text{C}$ 30 s, 63 $^{\circ}\text{C}$ 30 s, 72 $^{\circ}\text{C}$ 1 min, 35 个循环, 72 $^{\circ}\text{C}$ 10 min。将 3'-RACE 和 5'-RACE 所得产物送交上海生工生物工程有限公司测序。

表 1 *IGFBP* 序列克隆、荧光定量 PCR 及检测 SNP 位点的引物
Tab.1 Primers used for clone, quantitative real-time PCR and SNP test

引物名称 primer	引物序列 primer sequence(5'-3')	引物用途 usage
F1	TGTGTCGGGGAGAGGTTCCAGTG	中间片段扩增
R1	TGTAGCGGTGGAGCGAGCGTC	partial fragment amplification
UPM	CTAATACGACTCACTATAGGGCAAG CAGTGGTATCAACGCAGAGT	5'RACE
UPM (short)	CTAATACGACTCACTATAGGGC	5'RACE
P1	GCGACAGTTGTGGGTAAGGTGGGAG	5'RACE
Outer	TACCGTCGTTCCACTAGTGATT	3'RACE
P2	GGAGAGGTTCCAGTGTGCAC	3'RACE
F2	ATGAAATCGACGGCGTGT	DNA 全长克隆
R2	TCACTTCCTGCCTACAACCG	full-length clone of DNA
18SF	TCGGTCTATTGCGTTGGTTTT	荧光定量 PCR
18SR	CAGTTGGCATCG TTTATGGTCA	Real-time PCR
F3	GGGAGAGGTTCCAGTGTTC	荧光定量 PCR
R3	GCTTGGGAGAAACCTTGAC	Real-time PCR
SNP-F1	ACGTTGGATGGTACGCGCATTGTTCAATC	含 SNP 位点的片段扩增
SNP-R1	ACGTTGGATGTTGTTATCCGTTTCAACTG	SNP fragment amplification
SNP-P1	CATGTATATCTGAATTACGCAATATAAT	检测 SNP 位点 (SNP test probe)
SNP-F2	ACGTTGGATGTTGCGTTCATATGCACGG	含 SNP 位点的片段扩增
SNP-R2	ACGTTGGATGACGGTGTGTTTGTTCAAATGC	SNP fragment amplification
SNP-P2	TGGGCTAATTACGACGACCGTAGGA	检测 SNP 位点 (SNP test probe)
SNP-F3	ACGTTGGATGATCGTTGGTTAGGTGAATG	含 SNP 位点的片段扩增
SNP-R3	ACGTTGGATGCACTAAAAAAGCTAACCCAGAC	SNP fragment amplification
SNP-P3	CAGCTAACGACTACTGTACAT	检测 SNP 位点 (SNP test probe)
SNP-F4	ACGTTGGATGCGGTTTACATAATTCAGC	含 SNP 位点的片段扩增
SNP-R4	ACGTTGGATGTTACCAACGTGTACTGTCTC	SNP fragment amplification
SNP-P4	AAAGGGTACTGTCTCACATATAGGTA	检测 SNP 位点 (SNP test probe)

基因组序列的扩增方法是利用获得的全长 cDNA 序列, 根据 ORF 的前后 20 个碱基序列设计引物 F2 和 R2 扩增内含子, 以纯度高且完整的 DNA 为模板, 扩增体系为 DNA 2 μL , Ex Taq (5 U/ μL) 0.125 μL , 10 \times Ex Taq buffer (Mg^{2+} plus) 2.5 μL , dNTP Mixture (各 2.5 mmol/L) 2 μL , 上、下游引物 (10 $\mu\text{mol/L}$) 各 0.5 μL , RNase-free water 17.375 μL 。扩增条件为 94 $^{\circ}\text{C}$ 5 min, 94 $^{\circ}\text{C}$ 30 s, 50 $^{\circ}\text{C}$ 30 s, 72 $^{\circ}\text{C}$ 4 min, 35 个循

环, 72 $^{\circ}\text{C}$ 10 min。将产物送交上海生工生物工程有限公司测序。

1.6 荧光定量 PCR

用已获得的全长 cDNA 序列设计荧光定量 PCR 的引物, 用 7 个 5 倍浓度梯度稀释的 cDNA (5^0 、 5^{-1} 、 5^{-2} 、 5^{-3} 、 5^{-4} 、 5^{-5} 、 5^{-6}) 绘制标准曲线, 获得扩增效率介于 95% ~ 105% 且无非特异性扩增产物的引物, 以稀释 10 倍后的 cDNA 为扩增模板, 根据 SYBR[®] Premix Ex Taq II (TliRNaseH

Plus)试剂盒(TaKaRa)操作说明进行相对定量分析,荧光定量PCR扩增条件为95℃ 30 s,95℃ 10 s,57℃ 30 s,72℃ 30 s,40个循环,延伸阶段收集信号,从65℃到95℃,每个循环增加0.5℃,持续5 s获得解链温度,采集溶解曲线荧光信号。同时使用无酶水代替模板的反应体系做空白对照。

1.7 生长性状相关 SNP 位点筛选和 SNP 分型

在1龄缢蛏中随机选取6个体质量较大的个体(10.36±0.72)g和6个体质量较小的个体(5.42±0.58)g,用F2、R2引物扩增*IGFBP*基因DNA全长,反应体系和方法同基因组序列的扩增方法,将扩增产物送交上海生工生物工程有限公司测序。根据测序结果统计单核苷酸突变位点。SNP分型采用MassARRAY质谱检测方法,由上海邃志生物科技有限公司进行检测。具体方法为根据统计的单核苷酸突变位点进行PCR引物和延伸引物的设计,PCR引物用于扩增含SNP位点的片段,延伸引物用于延伸该检测SNP碱基。SNP分型操作第一步,通过多重PCR扩增目标序列,扩增体系为DNA 1 μL,HotstarTaq(5 U/μL) 0.1 μL,10×PCR buffer(含MgCl₂)1 μL,dNTP Mixture(各2.5 mmol/L)1 μL,多对PCR引物混合液(10 μmol/L)1 μL,RNase-free water 0.95 μL。扩增条件为94℃ 15 min,94℃ 20 s,56℃ 30 s,72℃ 1 min,45个循环,72℃ 3 min。第二步,对扩增后的反应液进行SAP处理,SAP反应液体系:RNase-free water 1.53 μL,10×SAP Buffer 0.17 μL,SAP酶(1 U/μL)0.3 μL。SAP处理程序:37℃ 40 min,85℃ 5 min,4℃保存。第三步,延伸反应,配制iPlex反应试剂:RNase-free water 0.755 μL,10×iPlex Buffer 0.2 μL,iPlex Termination mix 0.2 μL,多种延伸引物混合液0.804 μL,iPlex emzyme 0.041 μL,加入到上述反应液中,混匀。延伸程序:94℃ 30 s,94℃ 5 s,(52℃ 5 s,80℃ 5 s,5个循环),40个循环,72℃ 3 min,4℃保存。对延伸产物进行树脂纯化,然后进行Nanodispenser SpectroCHIP芯片点样,MassARRAY Analyzer Compac质谱检测,TYPER软件分析实验结果,从而获得分型数据。

1.8 数据统计与分析

实验数据采用平均值±标准差表示,实验结果通过SPSS 17.1软件进行SNP与生长性状的

关联分析,先对数据做单因素方差分析,若组间有显著差异,再进行Duncan's多重比较分析, $P < 0.05$ 时认为差异显著;并用Popgene软件对突变位点进行群体哈迪-温伯格平衡检验。采用DANMAN6.0软件将测序所得3'端和5'端的cDNA片段拼接得到cDNA全长,通过ORF Finder网站(<http://www.ncbi.nlm.nih.gov/projects/gorf/>)获得开放阅读框(open reading frame,ORF),利用NCBI网站(<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)进行氨基酸预测,用SignalP程序(<http://www.cbs.dtu.dk/services/SignalP/>)预测蛋白信号肽。利用ProtParam程序(<http://web.expasy.org/protparam/>)预测编码蛋白的理化性质。通过TMPred程序(http://www.ch.embnet.org/software/TMPRED_form.html)分析蛋白的跨膜结构。使用European Molecular Biology Laboratory(EMBL)的Interproscan程序(<http://www.ebi.ac.uk/interpro/scan.html>)预测蛋白结构域,使用ClustalW2程序(<http://www.ebi.ac.uk/Tools/msa/clustalw2/>)进行序列多重比对分析,MEGA 5.0和邻接法(Neighbor-Joining method,NJ)构建系统发育树;荧光定量PCR相关系数,解链温度等系列参数通过C1000™ Thermal cycler自带软件Bio-Red CFX Manager V1.6得出后导出至Microsoft Excel表格以供后续分析。目的基因与内参基因18S rRNA的相对表达量采用 $2^{-\Delta\Delta C_t}$ 方法计算,根据所得数据,采用SigmaPlot 12.0软件绘图。

2 结果

2.1 *IGFBP* 基因序列分析

经DANMAN软件拼接,缢蛏*IGFBP*基因cDNA全长共631 bp(GenBank登录号:KP245742)(图1),其5'端非编码区长60 bp,3'端非编码区长136 bp,开放阅读框435 bp(61~495 bp)。*IGFBP*的基因组DNA全长3122 bp(GenBank登录号:KP245743),包括1个内含子和2个外显子,内含子2687 bp,外显子分别为200和235 bp,剪切符合GT-AG剪接规则。

经ProParam推测,*IGFBP*基因编码蛋白含144个氨基酸,理论分子量为1.5 ku,等电点为6.62,呈弱酸性,SignalP的预测结果为,蛋白N端的1~18个氨基酸构成信号肽结构,经TMPred

初步推断,该蛋白是非跨膜蛋白。经 EMBL 网站 Service 工具 InterProScan 程序分析蛋白序列的保守结构域,此蛋白在 46 ~ 61 位氨基酸为 PS00222 (IGFBP-N-1)保守基序。在 21 ~ 97 位氨基酸为 IGFBP 保守结构域 (IPR000867) 的 3 个成员

PF00219 (IGFBP)、SM00121 (IB) 和 PS51323 (IGFBP-N-2),该结构域富含半胱氨酸残基。在 22 ~ 96 位氨基酸为 growth factor receptor domain 保守结构域 (SSF57184),该结构域为具有与 IGF 受体 (IGFR) 结合能力的保守基序。

```

1  ggagataaggactcttcaacacgaagtaatagacgcctctaataagacaactgccgacg
61  ATGAAATCGACGGCGTGTGTTTGTGCTGCTGTGTATGCTGGTGTGTCGGGGAGAGGTT
    M K S T A C F V L L C M L V L C R G E V
121 CCAGTGTGCACTGTGTAGGCTGCGACAATGTTACATGCCCGCCAGAGCCGACGAATTGC
    P V L H C V G C D N V T C P P E P T N C
181 TACCTAGATAGAATACCCTGCGGATGCTGTAATGTCTGTGCGCTAGGTCTAGGGAAAGG
    Y L D R I P C G C C N V C A L G L G E R
241 TGTCAAGGTTTCTCCGAAGCTGTGGCCCGGCCTGTGTGTCTGACCCCTGCAGGAACA
    C Q G F S A S C G P G L V C L T P A G T
301 TACGATGAGAGACCCCATGGTTCTCCGTATTCCACGGCATCTGCGTCTATGATTCCGGG
    Y D E R P P W F S V F H G I C V Y D S G
361 CACCATCATGCCGACCGACATTGGCTCCCACCTTACCCACAACGTGCGCWCTTGGGATC
    H H H A A P T L A P T L P T T V X L A I
421 GGGCACCAGGTTCCGACGCTCGCTCCACCGCTACATCCGGTTCCAACGGTTTCTCCGGTT
    G H Q V P T L A P P L H P V P T V S P V
481 GTAGGCAGGAAGTGAcgtcatgggaaaaaagtgacgttaattgattttatgacgtcatat
    V G R K *
541 gggaaaagggtgacaccatctgaattgcggtgacatcatttgattgtagatcaaaatac
601 agcgaatgattttatttgaaaaaaaaaaaaa

```

图 1 缢蛏 *IGFBP* 基因 cDNA 全长序列及其氨基酸序列

小写字母为非编码区,大写字母为编码区。框中依次为起始密码子,保守基序 PCGCCXXC 和 3'端加尾信号。下划线为信号肽区域,灰色阴影部分为 IGFBP-N 保守结构域,双下划线表示半胱氨酸残基,“*”表示终止密码子

Fig. 1 cDNA sequence and amino acid sequence of *IGFBP* in *S. constricta*

Lower case and upper case are used to indicate the 3'-and 5'-UTR and ORFs. Start codon (ATG), conserved motif (PCGCCXXC) and polyadenylation signal (aaaata) are marked by box. The signal peptide amino acid sequences are marked by black underline. The conserved N-terminal is shaded, and cysteine residues are annotated by double underline. The asterisk indicates the stop codon (TGA)

IGFBP 的 N 端结构高度保守,富含 12 个半胱氨酸的保守位点 (Cys-rich conserved site),脊椎动物的 GCGCCXXC 的保守基序,在缢蛏中表现为 PCGCCXXC (图 2)。同源性分析表明,缢蛏 IGFBP N 端与人和鼠的 IGFBP5 N 端相似度较其他结合蛋白的高,均为 40%。根据缢蛏、虾夷扇贝 IGFBP、杂色鲍 IGFBP7 与脊椎动物 IGFBP 的

氨基酸序列构建系统进化树 (图 3),对该进化树的拓扑结构进行了自展检验 (Bootstrap),重复抽样次数为 1 000 次。结果表明,系统进化树分为两大支,贝类 IGFBP 聚为一支,脊椎动物的聚为一支,这与生物进化的方向一致。脊椎动物的 IGFBP6 为一支,IGFBP1、2 和 4 聚为一支,IGFBP3 和 5 聚为一支。

<i>S.constricta</i>	1	MLVGGN--VTCPPEPTN--YDILP--GGENVLALGLGRKQGFSAAGPGVCLTPAGTYDEKPP--WFSVFIIGICV-----
<i>P.yessoensis</i> bp	1	FRCPNKS--PRPHLPDG--RVTLP--RQEEARRIGQTSAYSPLCAAGMEMNRGEALFVPTIMKTYIGIQNVVQ-----
<i>P.olivaceus</i> bp4	1	TRCPVSIERASRLP--DGGETVTEPGGGCPTALPKGAHGGVYSPGGTGLRQYPPG--VTRPLISLHMQGQVCTERMEEN--
<i>D.erio</i> bp1a	1	TRCAPSPERAECPAVDAGEEVLREPGGGCPTALPKGDPGGIYAPGGSLRCLPFG--EARPLHALTIGQAVCTESIPNH--
<i>D.erio</i> bp2a	1	RCPSIAERQAACQML--TTGGELYVEPGGGCPTALPKGDPGGIYAPGGSLRCLPFG--EARPLHALTIGQAVCTESIPNH--
<i>D.erio</i> bp3a	1	VRCEPCDGANVLRKPLPVDLDPVKEPGGGCPTALPKGDPGGIYAPGGSLRCLPFG--EARPLHALTIGQAVCTESIPNH--
<i>D.erio</i> bp5a	1	VRCEPCDGANVLRKPLPVDLDPVKEPGGGCPTALPKGDPGGIYAPGGSLRCLPFG--EARPLHALTIGQAVCTESIPNH--
<i>M.musculus</i> bp4	1	LRCPVSIERASRLP--DGGETVTEPGGGCPTALPKGAHGGVYSPGGTGLRQYPPG--VTRPLISLHMQGQVCTERMEEN--
<i>M.musculus</i> bp5	1	VRCEPCDGANVLRKPLPVDLDPVKEPGGGCPTALPKGDPGGIYAPGGSLRCLPFG--EARPLHALTIGQAVCTESIPNH--
<i>H.sapiens</i> bp3	1	VRCEPCDGANVLRKPLPVDLDPVKEPGGGCPTALPKGDPGGIYAPGGSLRCLPFG--EARPLHALTIGQAVCTESIPNH--
<i>H.sapiens</i> bp4	1	VRCEPCDGANVLRKPLPVDLDPVKEPGGGCPTALPKGDPGGIYAPGGSLRCLPFG--EARPLHALTIGQAVCTESIPNH--
<i>H.sapiens</i> bp5	1	VRCEPCDGANVLRKPLPVDLDPVKEPGGGCPTALPKGDPGGIYAPGGSLRCLPFG--EARPLHALTIGQAVCTESIPNH--

图2 缢蛭 IGFBP 的 N 端结构域与其他物种 IGFBP 的 N 端结构域的氨基酸序列比对
相同和相似氨基酸用相同颜色的阴影表示,其中暗红色 C 阴影为半胱氨酸残基

Fig. 2 N-terminal amino-acid sequences alignment of *S. constricta* with IGFBP genes from other species
Identical and similar residues are shaded in the same color, and cysteine residue are marked by the royal red color

表2 用于构建 NJ 树的氨基酸序列号
Tab. 2 Amino acid sequences used for NJ tree

基因名称 gene name	物种 species	GenBank 登录号 GenBank no.
IGFBP	虾夷扇贝 <i>P. yessoensis</i>	KF801670
IGFBP7	杂色鲍 <i>H. diversicolor</i>	JF501214. 1
IGFBP4	牙鲆 <i>Paralichthys olivaceus</i>	KJ000097
IGFBP1a	斑马鱼 <i>D. rerio</i>	NM-173283. 3
IGFBP2a	斑马鱼 <i>D. rerio</i>	NM-131458. 2
IGFBP3	斑马鱼 <i>D. rerio</i>	NM-205751. 2
IGFBP5	斑马鱼 <i>D. rerio</i>	KJ000097. 1
IGFBP6a	斑马鱼 <i>D. rerio</i>	NM-001161401. 1
IGFBP1	鸡 <i>G. gallus</i>	NM-001001294. 1
IGFBP2	鸡 <i>G. gallus</i>	NM-205359. 1
IGFBP4	鸡 <i>G. gallus</i>	NM-204353. 1
IGFBP5	鸡 <i>G. gallus</i>	XM-422069. 4
IGFBP1	人 <i>H. sapiens</i>	NM-000596. 2
IGFBP2	人 <i>H. sapiens</i>	NM-000597. 2
IGFBP3	人 <i>H. sapiens</i>	NM-001013398. 1
IGFBP4	人 <i>H. sapiens</i>	NM-001552. 2
IGFBP5	人 <i>H. sapiens</i>	CR4568809. 1
IGFBP6	人 <i>H. sapiens</i>	NM-002178. 2
IGFBP1	小鼠 <i>M. musculus</i>	NM-008341. 4
IGFBP2	小鼠 <i>M. musculus</i>	NM-008342. 3
IGFBP3	小鼠 <i>M. musculus</i>	NM-008343. 2
IGFBP4	小鼠 <i>M. musculus</i>	NM-010517. 3
IGFBP5	小鼠 <i>M. musculus</i>	NM-010518. 2
IGFBP6	小鼠 <i>M. musculus</i>	CT010184

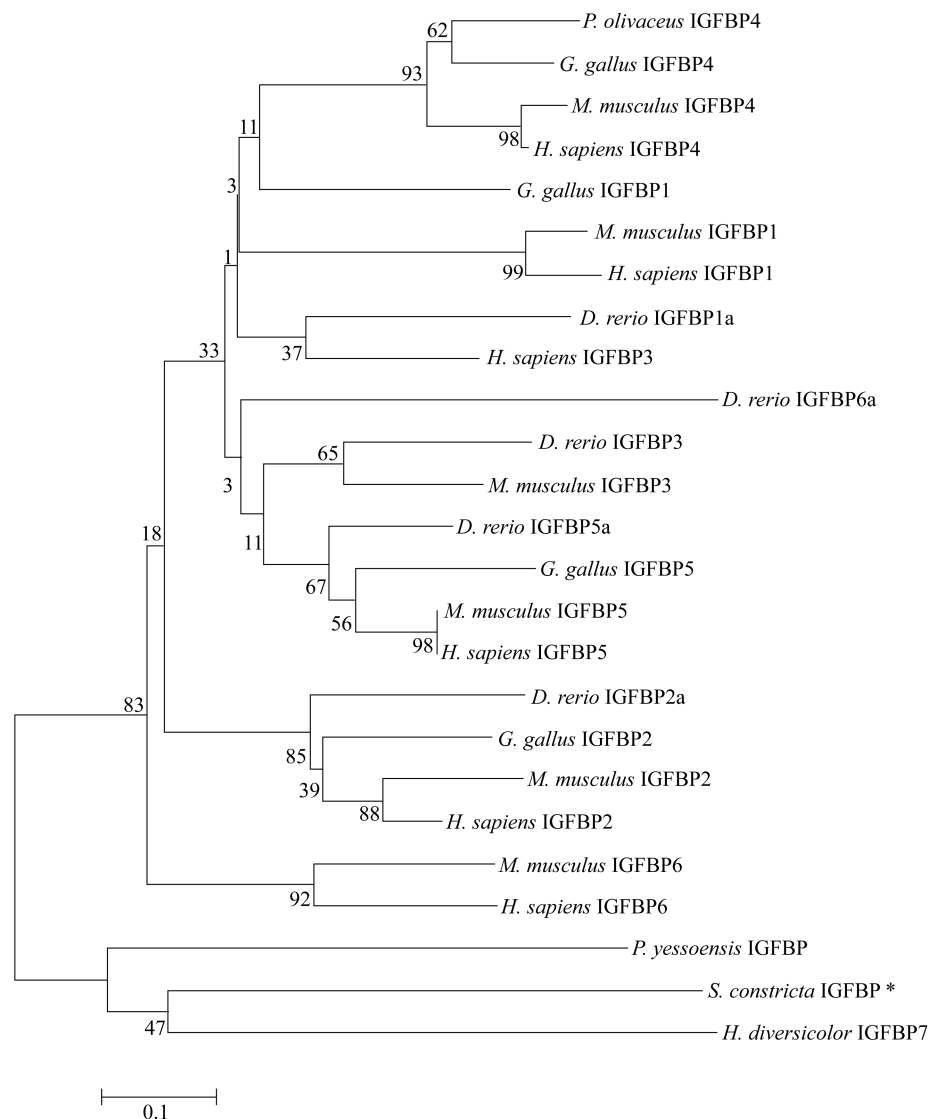


图3 基于 IGFBP 氨基酸序列构建的系统发育树

缢蛏 IGFBP 用“*”标注,建树物种 IGFBP 氨基酸序列号见表 2

Fig. 3 NJ phylogenetic tree based on IGFBP amino-acid sequences

The asterisk indicates the IGFBP of *S. constricta* and GenBank accession numbers are showed in Tab. 2

2.2 *IGFBP* 基因的时空表达分析

采用实时荧光定量 PCR 方法检测了 *IGFBP* mRNA 的相对表达情况,结果表明,*IGFBP* 在缢蛏成体的消化腺组织中表达水平最高,性腺次之,在外套膜、斧足、水管和鳃组织中只能检测到少量表达(图 4)。在缢蛏胚胎发育的卵裂期也可检测到 *IGFBP* 的表达,担轮期表达量下降,随着发育的进行,表达量逐渐升高,至单水管期,表达量明显上升,在稚贝期中的表达量达到最高(图 5)。

2.3 *IGFBP* 的 SNP 位点与生长性状的关联分析

在缢蛏 *IGFBP* 基因组 DNA 中,检测到 4 个

SNP 位点,分别位于基因组 DNA 的 1 216、1 665、2 168 和 2 420 bp 位置。对 240 个缢蛏个体进行 SNP 分型,发现 4 个 SNP 位点均发生 T/C 突变,2 168 bp 位置的突变检测到 3 种基因型,其他 3 个位置的突变检测到两种基因型(表 3)。4 个位点中,有 3 个位点符合哈迪-温伯格平衡。通过与生长性状的关联分析表明,2 168 位点的 TT 基因型个体的壳长大于 CT 基因型的缢蛏个体 ($P < 0.05$),TT 基因型个体的体质量显著大于 CT 和 CC 基因型的缢蛏个体 ($P < 0.05$)。*IGFBP* 基因组的其他 3 个位点的突变与生长性状无显著相关性。

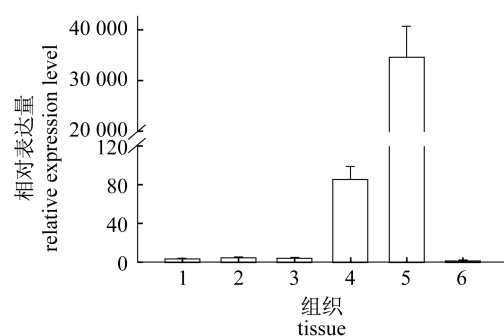


图4 缢蛏 *IGFBP* 组织分布图

1. 外套膜; 2. 鳃; 3. 斧足; 4. 性腺; 5. 消化腺; 6. 水管

Fig. 4 Expressions of *IGFBP* gene in six tissues of *S. constricta*

1. mantle; 2. gill; 3. foot; 4. gonad; 5. digestive gland; 6. siphon

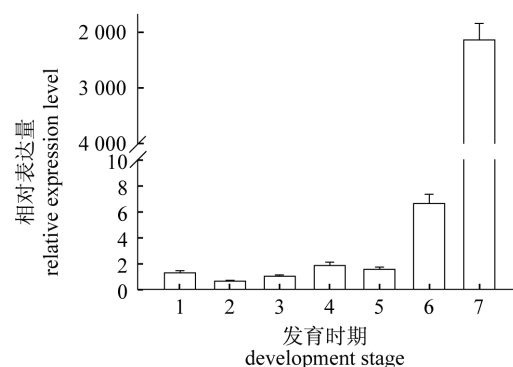


图5 缢蛏 *IGFBP* 发育时期表达图

1. 卵裂期; 2. 担轮期; 3. 面盘期; 4. 壳顶期; 5. 匍匐期; 6. 单水管期; 7. 稚贝期

Fig. 5 Relative mRNA expressions of *IGFBP* gene during early development *S. constricta*

1. embryogenesis; 2. trochophore; 3. veliger; 4. umbo larvae; 5. creeping larvae; 6. single pipe stage; 7. juvenile

表3 缢蛏 *IGFBP* 的 SNP 位点与生长性状的关联分析

Tab. 3 Growth traits of clams with different genotypes and association analysis

位置 locus	基因型 genotype	样本量/F N/F	哈迪-温伯格 平衡(P值) HWE(P value)	壳长/mm SL	壳宽/mm SW	壳高/mm SH	体质量/g BW
d. 1216. C. T	TC	13/0.06	$P = 0.91$	52.17 ± 3.39	12.76 ± 1.01	17.36 ± 0.94	8.31 ± 1.73
	CC	220/0.94		53.01 ± 2.12	12.36 ± 0.90	17.55 ± 0.92	8.10 ± 1.21
d. 1665. C. T	CT	2/0.01	$P = 0.99$	52.05 ± 0.95	12.82 ± 1.23	18.02 ± 0.41	8.25 ± 1.77
	TT	235/0.99		52.99 ± 2.24	12.39 ± 0.90	17.54 ± 0.92	8.13 ± 1.24
d. 2168. C. T	CC	208/0.937	$P = 0.00$	52.98 ± 2.13^{ab}	12.38 ± 0.90	17.55 ± 0.89	8.09 ± 1.21^b
	CT	8/0.036		51.51 ± 2.57^b	12.35 ± 1.20	16.97 ± 0.77	7.63 ± 1.60^b
	TT	6/0.027		54.02 ± 3.85^a	12.89 ± 0.40	17.68 ± 0.82	9.25 ± 1.44^a
d. 2420. C. T	CT	3/0.01	$P = 0.99$	54.70 ± 1.40	12.62 ± 0.81	17.99 ± 0.95	8.50 ± 1.73
	TT	230/0.99		52.97 ± 2.23	12.40 ± 0.91	17.56 ± 0.90	8.14 ± 1.24

注: N 表示基因型样本量, F 表示基因型频率, HWE 表示哈迪-温伯格平衡, SL 表示壳长 (mm), SH 表示壳高 (mm), SW 表示壳宽 (mm), BW 表示体质量 (g), 数值均以平均值 \pm 标准差表示。a 与 b 有显著性差异 ($P < 0.05$); ab 与 a、b 均无显著性差异

Notes: N, number of clams; F, genotype frequency; HWE means Hardy-Weinberg Equilibrium; SL, shell length (mm); SH, shell height (mm); SW, shell width (mm); BW, body weight (g). The growth traits are given as the mean \pm standard deviation. Within each column of each locus, the value with superscript a is significantly different from that with superscript b ($P < 0.05$), and value with superscript ab is not significantly different from that with a or with b

3 讨论

分泌蛋白是在细胞内合成,在胞外发挥生物学作用的蛋白。信号肽是指导蛋白质跨膜转移的氨基酸序列,一般位于分泌蛋白的 N 端。本实验克隆的缢蛏 *IGFBP* 基因编码的氨基酸经 SignalP 程序分析, N 端的前 18 个氨基酸为信号肽结构。经 TMpred 程序分析为非跨膜蛋白,说明该蛋白属于分泌型蛋白。此推测与已知的 *IGFBP* 的作用方式相符^[12]。*IGFBP* 分泌到细胞外,与 IGF 结

合并延长 IGF 的半衰期,促进或抑制 IGF 的生物活性^[13]。

蛋白质的进化在生物进化的过程中起中心作用,类似结构域出现在不同的蛋白质分子中,推测其可能是来源于共同祖先的基因^[14]。脊椎动物中的 *IGFBP* 有高度同源性,其共同的结构特征: 1 个含 12 个 Cys 残基的高度保守的 N-末端区、1 个含 6 个 Cys 残基的 C-末端保守区和 1 个高度可变的 L 区^[15]。在脊椎动物中, Cys 残基形成系列二硫键, N 端主要与 IGF 结合,并使 IGF 的部分

氨基酸残基暴露,以便与 IGF-IR 结合^[15-16]。C 端与 IGF 或者 N 端连接,保持 N 端与 IGF 的结合和稳定^[16]。L 区是每个 IGFBP 最独特的区域,它使 IGFBP 的结构和功能有差异^[16]。缢蛏 IGFBP 成熟肽的 N 端保守区包含 12 个 Cys 氨基酸,与脊椎动物相同,由此推测缢蛏 IGFBP 的 N 端有与脊椎动物相似的功能,能与 IGF 配体结合,并调节其活性^[15]。缢蛏 IGFBP N 端的 PCGCCXXC 保守基序与脊椎动物的 GCGCCXXC 稍有差异,且 IGFBP 的 L 区大部分缺失,C 端不含甲状腺球蛋白结构域(thyroglobulin type-1 repeat)^[17]。缢蛏的 IGFBP 结构与虾夷扇贝的相似^[10]。贝类 IGFBP 与脊椎动物 IGFBP 在结构上的差异暗示两者之间存在不同的生物学功能。这可能与 IGFBP 蛋白的进化相关。结构域可能是蛋白质进化的一种遗传单位,产生新蛋白的方式之一是结构域的组合^[18]。推测 IGFBP 的祖先基因可能只有 N 端保守结构域,在进化的过程中,与甲状腺球蛋白结构域进行组合,产生新的 IGFBP 蛋白,由于两个结构域的折叠需要更大的空间,所以 L 区变长,随着生物的进化,IGFBP 趋向于专一性和特异性,因此分化出了 IGFBP1~6。IGFBP 在脊椎动物中经过长期的进化和演变,分枝出 IGFBP1、3 和 6,随后的基因复制事件使 IGFBP1 分化出 IGFBP2 和 4,IGFBP3 分化出 IGFBP5^[19]。

目前,关于贝类 IGFBP 研究很少。仅在杂色鲍^[9]和虾夷扇贝^[10]中有相关报道。在本实验中,对 *IGFBP* 在缢蛏外套膜、水管、斧足、鳃、性腺和消化腺 6 个组织中的实时荧光定量分析表明,*IGFBP* 基因在消化腺中表达量高,在性腺中次之,在外套膜等其他组织中表达量低。推测其与消化腺的功能相关,缢蛏消化腺是合成、分泌消化酶的主要器官,具有很强的细胞吸收和细胞内消化的功能,还参与无机盐的代谢^[20],与脊椎动物中肝脏的功能类似。脊椎动物中,IGF 有自分泌和旁分泌的功能,但主要在肝脏中合成,并运输到其他组织中发挥作用^[2]。因此,推测缢蛏 IGFBP 作为 IGF 的运输蛋白,主要在消化腺中合成。脊椎动物中肝脏是分泌 IGFBP1 的主要器官^[21]。研究表明,成体斑马鱼的 IGFBP1 只在肝脏中表达^[22]。五条鲷(*Seriola quinqueradiata*)^[23]和牙鲆的 IGFBP 在肝脏中表达量最高^[24]。此外,IGFBP1 在牙鲆胚胎发育的各阶段的表达量均处

于较低水平,仅在牙鲆出膜后 3 d 及变态前出现高峰,在变态期表达量最高,对牙鲆的变态发育有重要作用^[24]。缢蛏 IGFBP 从受精卵至匍匐期阶段,表达量均处于较低水平,单水管期表达水平明显升高,稚贝期表达水平最高,表明 IGFBP 可能对缢蛏的发育有重要作用。

缢蛏 *IGFBP* 的基因组结构包含 1 个内含子和两个外显子。与生长性状的关联分析表明,2 168 位点的 TT 基因型的缢蛏体质量和壳长显著大于 TC 基因型。在虾夷扇贝的 *IGFBP* 中,也检测到了 1 个与体质量和壳长显著相关的突变位点^[10]。IGFBP 在脊椎动物中主要与 IGF 结合调控生长和发育^[3,5]。而且内含子位点的突变与基因转录调控有关^[25]。因此,基因的位点突变有可能会影响到缢蛏 IGFBP 的翻译、结构和功能,进而对生长性状产生影响^[25]。*IGFBP* 的位点突变对缢蛏生长性状的影响机制还有待进一步的研究。

参考文献:

- [1] Xu F S, Zhang S P. Illustrated Guide to Marine Bivalve of China[M]. Beijing: Science Press, 2008: 211-213. [徐凤山,张素萍. 中国海产双壳贝图志. 北京:科学出版社,2008:211-213.]
- [2] Wood A W, Duan C M, Bern H A. Insulin-like growth factor signaling in fish [J]. International Review of Cytology, 2005, 243: 215-285.
- [3] Duan C M, Xu Q J. Roles of insulin-like growth factor (IGF) binding proteins in regulating IGF actions [J]. General and Comparative Endocrinology, 2005, 142(1-2): 44-52.
- [4] Moriyama S, Ayson F G, Kawauchi H. Growth regulation by insulin-like growth factor-I in fish [J]. Bioscience Biotechnology and Biochemistry, 2000, 64(8): 1553-1562.
- [5] Duan C M. Specifying the cellular responses to IGF signals; Roles of IGF-binding proteins [J]. Journal of Endocrinology, 2002, 175(1): 41-54.
- [6] Zhang L J, He M X. Effects of growth hormone and insulin-like growth factor-I on shell growth related gene expression in *Pinctada fucata* Gould [J]. Journal of Tropical Oceanography, 2012, 31(2): 96-101. [张立娟,何毛贤. 生长激素与胰岛素样生长因子-I 对马氏珠母贝壳生长形成相关基因表达的影响. 热带海洋学报, 2012, 31(2): 96-101.]
- [7] Shi Y, Guan Y Y, He M X. Molecular identification

- of insulin-related peptide receptor and its potential role in regulating development in *Pinctada fucata* [J]. *Aquaculture*, 2013, 408409: 118 – 127.
- [8] Gricourt L, Bonnac G, Boujard D, *et al.* Insulin-like system and growth regulation in the Pacific oyster *Crassostrea gigas*: hrIGF-1 effect on protein synthesis of mantle edge cells and expression of an homologous insulin receptor-related receptor [J]. *General and Comparative Endocrinology*, 2003, 134 (1): 44 – 56.
- [9] Li N, Zhang Z P, Zhang L L, *et al.* Insulin-like growth factor binding protein 7, a member of insulin-like growth factor signal pathway, involved in immune response of small abalone *Haliotis diversicolor* [J]. *Fish and Shellfish Immunology*, 2012, 33 (2): 229 – 242.
- [10] Feng L Y, Li X, Yu Q, *et al.* A Scallop IGF Binding Protein Gene: Molecular Characterization and Association of Variants with Growth Traits [J]. *PLoS One*, 2014, 9 (2): e0089039.
- [11] Feng B B, Dong L L, Niu D H, *et al.* Identification of Immune Genes of the Agamaki Clam (*Sinonovacula constricta*) by Sequencing and Bioinformatic Analysis of ESTs [J]. *Marine Biotechnology*, 2010, 12 (3): 282 – 291.
- [12] Kelley K M, Oh Y, Gargosky S E, *et al.* Insulin-like growth factor-binding proteins (IGFBPs) and their regulatory dynamics [J]. *International Journal of Biochemistry and Cell Biology*, 1996, 28 (6): 619 – 637.
- [13] Jones J I, Clemmons D R. Insulin-like growth factors and their binding proteins: Biological actions [J]. *Endocrine Reviews*, 1995, 16 (1): 3 – 34.
- [14] Xie X Y, Li X, Cao C. Implications of protein domain combination network in eukaryotes [J]. *Acta Biophysica Sinica*, 2010, 12 (26): 1145 – 1153. [谢雪英, 李鑫, 曹晨. 基于复杂网络的蛋白质结构域组进化分析. *生物物理学报*, 2010, 12 (26): 1145 – 1153.]
- [15] Hwa V, Oh Y, Rosenfeld R G. The insulin-like growth factor-binding protein (IGFBP) superfamily [J]. *Endocrine Reviews*, 1999, 20 (6): 761 – 787.
- [16] Schneider M R, Wolf E, Hoeflich A, *et al.* IGF-binding protein-5: Flexible player in the IGF system and effector on its own [J]. *Journal of Endocrinology*, 2002, 172 (3): 423 – 440.
- [17] Novinec M, Kordis D, Turk V, *et al.* Diversity and evolution of the thyroglobulin type-1 domain superfamily [J]. *Molecular Biology and Evolution*, 2006, 23 (4): 744 – 755.
- [18] Lian D J, Xu G J. Structure domains in protein structure and function [J]. *Progress in Biochemistry and Biophysics*, 1997, 24 (6): 482 – 486. [廉德君, 许根俊. 蛋白质结构与功能中的结构域. *生物化学与生物物理进展*, 1997, 24 (6): 482 – 486.]
- [19] Rodgers B D, Roalson E H, Thompson C. Phylogenetic analysis of the insulin-like growth factor binding protein (IGFBP) and IGFBP-related protein gene families [J]. *General and Comparative Endocrinology*, 2008, 155 (1): 201 – 207.
- [20] Xu X H, Zheng J S. Histological and histochemical studies on digestive diverticula of *Sinonovacula constricta* [J]. *Journal of Fisheries of China*, 2003, 27 (3): 225 – 230. [许星鸿, 郑家声. 缢蛭消化盲囊的组织学与组织化学. *水产学报*, 2003, 27 (3): 225 – 230.]
- [21] Lee P D, Conover C A, Powell D R. Regulation and function of insulin-like growth factor-binding protein-1 [J]. *Proceedings of the Society for Experimental Biology and Medicine*, 1993, 204 (1): 4 – 29.
- [22] Maures T J, Duan C M. Structure, developmental expression, and physiological regulation of zebrafish IGF binding protein-1 [J]. *Endocrinology*, 2002, 143 (7): 2722 – 2731.
- [23] Pedroso F L, Fukada H, Masumoto T. Molecular characterization, tissue distribution patterns and nutritional regulation of IGFBP-1, -2, -3 and -5 in yellowtail, *Seriola quinqueradiata* [J]. *General and Comparative Endocrinology*, 2009, 161 (3): 344 – 353.
- [24] Zhai W Y. Identification and Functional Analysis of IGFBP-1 and HIF-1 α of *Paralichthys olivaceus* [D]. Shanghai: Shanghai Ocean University, 2012. [翟万营. 牙鲆 IGFBP-1、HIF-1 α 基因的克隆及功能研究. 上海: 上海海洋大学, 2012.]
- [25] John S M. Introns: Evolution and function [J]. *Current Opinion in Genetics and Development*, 1994, 4 (6): 823 – 831.

Molecular characterization of IGFBP and association analysis with growth traits in the razor clam *Sinonovacula constricta*

XIE Shumei, NIU Donghong^{*}, RUAN Haideng, WANG Ze, WANG Fei, CHEN Shi'e, LI Jiale

(Key Laboratory of Freshwater Aquatic Genetic Resource, Ministry of Agriculture,
Shanghai Ocean University, Shanghai 201306, China)

Abstract: Insulin like growth factor binding protein (IGFBP) is a member of IGFBP family participating in transportation, localization and biological regulation of IGFs. Here, we conducted RACE-PCR and LONG-PCR for the cloning of full length cDNA and DNA of IGFBP gene in *Sinonovacula constricta*. Quantitative real-time PCR analysis revealed that IGFBP transcript was expressed in a wide range of tissues and developmental stages. Additional growth-linked SNP were discovered based on association study. The full length of IGFBP cDNA sequence consists of 631 base pairs (bp), comprising a 60 bp 5' untranslated region (UTR), a 136-bp 3' UTR, and an open reading frame (ORF) which is 435 bp in length. The ORF encodes 144 amino acids with a signal peptide (18 amino acids) and 12 cysteine residues in the conserve N-terminal. There is one intron (2 687 bp) and two exons (200 and 245 bp) in the sequence of IGFBP gene (3 122 bp). Based on quantitative real-time PCR analysis, we found the highest level of IGFBP expression in the digestive gland. Also the transcript was highly detected in juvenile stages rather than other stages. Four SNPs were identified in IGFBP genomic DNA and one of them was significantly associated with the shell length and body weight. Our results indicated that IGFBP gene plays important roles in the growth and development in *S. constricta*.

Key words: *Sinonovacula constricta*; insulin-like growth factor binding protein; gene expression; SNP; association analysis

Corresponding author: NIU Donghong. E-mail: dhniu@shou.edu.cn